Definition and calibration of the hadronic recoil in view of the measurement of the $W$ boson mass with the CMS experiment

Candidato
Nicolò Foppiani

Relatore
Prof. Luigi Rolandi

Anno Accademico 2016/2017
Abstract

Although the Standard Model of particle physics might appear as a complete theory, there are several theoretical problems and experimental observations which suggest that it is not the complete and definitive theory. For this reason, new physics is searched for both directly, and through precise measurements of standard model observables.

A measurement of the mass of the $W$ boson with a relative precision of $10^{-4}$ is of primary interest because of the small uncertainty in its theoretical prediction, and because the LHC experiments have already collected several hundreds of millions of $W$ boson decays, which allow precise measurements of its properties. Such a measurement may represent a turning point in the world of particle physics, showing that the SM is inconsistent at this energy scale.

In order to achieve such precision, it is requested to master the detector, the analysis, and the theoretical predictions at an unprecedented level. A study of the hadronic recoil is described in this thesis, resulting in a new and better experimental definition of this quantity. The power of the new definition is tested in terms of systematic uncertainties on the $W$ mass measurement, which are evaluated in a new and original way.
# Contents

## Introduction vii

1 The Standard Model of particle physics 1
   1.1 Matter and interactions .......................... 1
   1.2 The origin of the mass: the Higgs mechanism .......... 4
   1.3 Electroweak precision tests ........................ 7
   1.4 Motivation for a precise W mass measurement ......... 10
   1.5 Conclusions ...................................... 12

2 The CMS experiment at the LHC 13
   2.1 The Large Hadron Collider .......................... 13
   2.2 The Compact Muon Solenoid experiment ................. 16
      2.2.1 The superconductive magnet ....................... 19
      2.2.2 The silicon tracker ............................. 19
      2.2.3 The electromagnetic calorimeter ................. 21
      2.2.4 The Hadron Calorimeter .......................... 23
      2.2.5 The muon system ............................... 24
      2.2.6 The particle flow reconstruction algorithm ....... 24
   2.3 Conclusions ...................................... 26

3 Measuring the W mass 27
   3.1 A historical overview ................................ 27
      3.1.1 W mass measurements at LEP II ................... 28
      3.1.2 W mass measurements at Tevatron ................. 30
      3.1.3 The first attempts at the LHC: ATLAS and CMS .... 31
   3.2 The W mass measurement at hadron colliders .......... 32
      3.2.1 Lepton transverse momentum and transverse mass .... 34
CONTENTS

3.2.2 Systematic uncertainties ........................................ 36
3.2.3 Z events to calibrate and validate the analysis ............... 43
3.3 Conclusions .......................................................... 43

4 Study of $W$ production at LHC on simulated events .............. 45
4.1 The $W$ production mechanism ..................................... 45
4.1.1 The origin of the $W$-$p_T$ ...................................... 47
4.1.2 The rapidity spectrum ............................................ 48
4.2 A closer look at the recoil .......................................... 49
4.3 Event selection ....................................................... 51
4.3.1 Effect of the selection on the transverse mass ............... 51
4.3.2 Selecting the events on the data-sets .......................... 51
4.4 Impact of the experimental resolution on the transverse mass . 53
4.4.1 The impact of the lepton on the transverse mass ............. 53
4.4.2 The problem of the recoil definition ........................... 54
4.4.3 The impact of the recoil on the transverse mass ............. 55
4.4.4 Statistical sensitivity in bins of the lepton $p_T$ ............. 55
4.5 Conclusions .......................................................... 57

5 Towards a better recoil definition ................................. 59
5.1 Defining the recoil .................................................. 59
5.1.1 Traditional definitions .......................................... 61
5.1.2 Recoil resolution ................................................ 63
5.2 Improving the recoil definition .................................... 66
5.2.1 Semi-parametric regression ................................... 67
5.2.2 Deep neural networks .......................................... 69
5.2.3 A semi-parametric regression for the recoil ................. 71
5.2.4 Possible parametrizations of the correction ................. 73
5.2.5 Evaluating the goodness of a definition ....................... 73
5.3 $c_{1,2}$ regression ................................................... 74
5.3.1 $c_{1,2}$ distributions ............................................ 74
5.3.2 Input variables ................................................ 77
5.3.3 Picking a recoil estimator from the distribution ............ 79
5.3.4 Closure plots .................................................... 79
Introduction

Particle physics is now facing deep questions. On the one hand the Standard Model of particle physics, the theory which summarizes all the knowledge on particle physics, can predict the results of all the experiments we are doing in our laboratories with great precision. On the other hand, there are several theoretical questions and experimental observations which suggest that it may only be a low energy approximation of a more general theory.

The Large Hadron Collider, the largest particle accelerator ever built, is delivering an incredibly large amount of collisions at the highest center of mass energy ever reached. This environment allow direct searches of new physics phenomena at high energy, however, up to now, no positive result has been obtained. Together with direct searches, precise measurements of the electroweak observables represent a fundamental tool to test the consistency of the standard model, and to possibly unveil quantum effects brought by new physics at larger energy scale. Such tests have already shown their power in predicting precisely the value of the masses of the Top Quark and of the Higgs Boson before their discovery.

Among the various electroweak observables, the mass of the $W$ boson is of major interest. This is firstly motivated by the good precision with which the Standard Model predicts this quantity, $M_{W}^{\text{theo}} = 80.3618 \pm 0.0080$ MeV [1], in comparison with the world average of the $W$ boson mass measurements, which has an uncertainty of 15 MeV [2]. In order to exploit the good theoretical prediction, a measurement of the mass of the $W$ boson at the level of $10^{-4}$ is needed. This kind of measurement can now be achieved at the LHC, because of the large amount of $W$ boson decays already collected by the LHC experiments. Such a measurement can represent a turning point the world of particle physics, showing that the SM is inconsistent at this energy scale. This would imply that the standard model as we know it is a low energy approximation of a new and undiscovered theory.

The CMS experiment has already made public a preliminary work, where the $Z$ mass has been measured with the $W$ boson analysis, and it is now planning to deliver a high precision measurement of the $W$ boson mass in the next few years. Such measurement requires an unprecedented understanding and control of the detector, the analysis, and the theoretical calculations. The analysis is performed by looking at events where the $W$ boson decays into a lepton and a neutrino. The measurement is performed through a simulated template fit to the distributions of variables which are sensitive to $M_{W}$, such as the lepton transverse momentum and the transverse mass. This measurement heavily relies on the MC simulation: each discrepancy between
the simulation and the data manifests as a systematic uncertainty.

Thanks to the large collected dataset, the main goal of the analysis is to reduce the systematic uncertainty. One of the biggest uncertainties at the moment is represented by the theoretical prediction of the $W$-$p_T$ spectrum. The most advanced calculations (NNLO+NLLO) predict it with large uncertainties, at the 10% level, whereas the latest experimental result, made public by the ATLAS collaboration, shows an inconsistency between the calculations and the data, preferring the simpler calculations (LO). As this uncertainty is expected to be complex and large if properly assessed, an experimental method to mitigate this effect is represented by the measurement of the transverse mass. It requires the measurement of two experimental objects: the lepton and the hadronic recoil. The lepton resolution is very good, and the aim is on the correct calibration of the scale of its transverse momentum. The recoil has instead a very poor experimental resolution, but its importance is highlighted by the fact that a better and better experimental definition of the recoil result in a transverse mass which is less and less dependent on the $W$-$p_T$ spectrum. Therefore, finding a better experimental definition of the recoil is one of the major tasks in order to reduce the systematic uncertainty related to the uncertainty on the prediction of the $W$-$p_T$ spectrum.

This thesis is divided into two parts. The first three chapters provide an introduction to the Standard Model of particle physics, the experimental apparatus used for the analysis, and the history of the $M_W$ measurement, focusing particularly on the LHC measurement and on the problem faced in order to achieve a measurement with a $10^{-4}$ precision. The second part contains the original work produced during this thesis. Chapter 4 shows a careful study of the hadronic recoil, highlighting its production mechanism and the experimental issues in its measurement. Chapter 5 illustrates the guidelines for a new and better definition of the hadronic recoil, studying the concepts of the recoil resolution and the effect on the transverse mass. These concepts lead to the development of a new technique based on machine learning algorithms, in order to provide a new definition of the recoil which exploits the information collected by the detector in a much better way compared to the previous definitions. In chapter 6 the power of different recoil definitions is studied in terms of systematic uncertainties. A factorization approach has been introduced in order to treat uncorrelated uncertainties, and a toy version of the fit, which allows for the study of systematic uncertainties separately from each other, has been developed. Using these tools, a ranking of the various recoil definitions in terms of systematic uncertainty due to an imperfect knowledge of the $W$-$p_T$ spectrum versus the expected statistical uncertainty is obtained. Eventually, the systematic uncertainty due to a mismodeling of the recoil is studied. A method for deriving specific corrections for the MC is developed, and its power is assessed for the different recoil definitions.

All the original work illustrated in the course of this thesis has been performed by the candidate together with Olmo Cerri, another Master student. The two students have worked closely together, resulting in the fact that the work made by one or the other is largely indistinguishable. If this work is cited in a scientific context, please consider also citing the corresponding thesis written by Olmo Cerri [3].
Chapter 1

The Standard Model of particle physics

The Standard Model (SM) of particle physics is the theory which summarizes all our knowledge on particle physics, describing the elementary particles that form matter as well as their interactions, from a quantum mechanical point of view. It describes everything that is observed and reproduced in our laboratories. All the elementary particles predicted by the standard model have been observed in high energy experiments, but we know that this is not the end of the story: there are several experimental observations and theoretical questions that cannot be addressed within the SM, such as the dark matter evidence, the acceleration of the universe, the hierarchy problem and the baryon asymmetry problem. Furthermore there are theoretically reasons to think that the SM, as we know it, cannot describe physics at all the physics scales, as it will be discussed later. The discovery of new physics beyond the SM might answer some of these questions: however, today there is no clear hint of the direction where to look for something new.

The aim of this chapter is to provide a general overview of the basic concepts of the SM, which will serve as a theoretical background of the main subject of this thesis. It focuses particularly on the main theoretical aspects of the SM and on its predictive power through precision tests. This eventually leads to the motivation for a more precise measurement of the mass of the W boson. A more complete overview of this topic is beyond the purpose of this thesis, and can be found in Ref. [4] and Ref. [5].

In the typical situations in which the particle physics experiments are performed, the gravitational interaction is negligible with respect to the other interactions. Therefore a discussion about gravity is out of the scope of this thesis; a complete summary of the description of the gravitational interaction can be found in Ref. [6].

1.1 Matter and interactions

The Standard Model is a non abelian gauge theory that describes all the elementary particles that form matter as well as their interactions. The constituents of matter are fermions of spin 1/2, whereas the interactions are mediated by spin 1 bosons. The interactions are fixed once the
Chapter 1. The Standard Model of particle physics

gauge symmetry groups are assigned. These symmetry groups refer to global transformations, that give conserved currents which the gauge fields are coupled to. Since it is a gauge theory, the Lagrangian is also invariant under local gauge transformations. These transformations are not physical and they express only the redundancy of the description of the mass-less gauge fields with Lorentz 4-vectors. The gauge symmetry groups for the SM are:

\[ SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \]

\( SU(3)_C \), where C stands for color, is the group of Quantum Chromodynamics (QCD) and describes the so called strong interactions. \( SU(2)_L \otimes U(1)_Y \) instead describes the electroweak interactions (EW), where the electric charge \( Q \), whose origin will be better discussed in the next chapter, can be expressed in terms of the weak isospin \( T_3 \) and of the hypercharge \( Y \) using the formula: \( Q = \frac{1}{2}(Y + T_3) \).

\[ Q = \frac{1}{2}(Y + T_3) \]

Figure 1.1: The particles of the standard model: in purple the quarks are shown, in green it is possible to see the leptons, whereas the gauge bosons appear in the red boxes.

The fermions are split in two categories: quarks and leptons. Quarks interact both via electroweak and strong interactions. There are two kind of quarks: the up type and the down type. They differ for the hypercharge and for the weak isospin, that reflect in a difference in the electric charge: +2/3 for the up-type and -1/3 for the down-type. Leptons interact only via
electroweak interaction. There are two kind of leptons: the charged ones, and the neutrinos. Their hypercharge and weak isospin make the charged lepton with electric charge -1 and the neutrino with no electric charge. The left-handed fermions are arranged in SU(2)\(_L\) doublets, containing, for the quarks, the up and down type left-handed quarks, and for the leptons, the left-handed charged lepton and the neutrino, whereas the right-handed ones are singlets under SU(2)\(_L\) transformations. There are three generations of fermions: each of them contains an up-type and a down-type quark, a charged lepton and a neutrino. The three generations are equal in terms of quantum numbers but differ for the masses. Only the lighter generation is stable: the other particles decay into lighter ones via the EW interaction. The first generation contains the constituents of ordinary stable matter: the up and down quarks form protons and neutrons, that, together with the charged lepton, better known as the electron, represent the building blocks of the atoms.

There are eight gluons, that mediate the strong interaction SU(3)\(_C\), three bosons for SU(2)\(_L\) and one boson for U(1)\(_Y\). Since SU(3)\(_C\) and SU(2)\(_L\) are non-abelian groups, the mediators are charged under their own gauge group, whereas this is not the case for U(1)\(_Y\).

The Lagrangian of the SM contains all the terms that are Lorentz-invariant, Gauge-invariant and renormalizable, meaning that divergences that arise computing loop diagrams can be handled with a finite number of counter-terms

\[
\mathcal{L} = -\frac{1}{4g_1^2} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4g_2^2} \sum_{j=1}^{3} W^i_{\mu\nu} W^{i\mu\nu,j} - \frac{1}{4g_3^2} \sum_{\alpha=1}^{8} G^\alpha_{\mu\nu} G^{\mu\nu,\alpha} + \sum_{f=\{Q_L,u_R,d_R,L_L,e_R\}} \sum_{i=1}^{3} i \bar{\psi}_f^i D \psi_f^i \tag{1.1}
\]

where \(B_{\mu\nu}, W^i_{\mu\nu}, \) and \(G^\alpha_{\mu\nu}\) are the field strengths\(^1\) related to the hypercharge, weak, and strong interactions, respectively, with relative coupling constants \(g_1, g_2,\) and \(g_3.\) The index \(j\) goes on the three bosons of the weak SU(2)\(_L\) group, whereas \(\alpha\) labels the eight gluons. The fermionic fields are represented with \(\psi_f^i,\) where \(f\) labels the various fermions \(\{Q_L, u_R, d_R, L_L, e_R\},\) where the capital letter refers to the doublets under SU(2)\(_L\) and the small letter to the singlets, and \(i\) labels the three fermionic generations. There is no right-handed neutrino in this formulation. Although the fact that neutrinos are massive is well assessed from the flavour oscillation experiments, they are not present in this formulation of the SM. They can be implemented in the SM in two ways: the Dirac way introducing a right-handed neutrino and a Yukawa coupling as for the other fermions, or a Majorana way, without any right-handed neutrino, but introducing a non-renormalizable term. The first line of the Lagrangian contains the kinetic terms for the gauge fields, whereas the second one contains the kinetic term for the fermions, where \(D\) stands for the covariant derivative:

\[
D_\mu = \partial_\mu - i \frac{g_1}{2} Y B_\mu - i \frac{g_2}{2} T_j W^j_\mu - i \frac{g_3}{2} S_\alpha G^{\alpha}_\mu \tag{1.3}
\]

\(^1\)The field strength is the tensor built from the gauge field \(A_\mu\) in the following way: \(F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.\) In the case of the electromagnetism, the \(F_{\mu\nu}\) contains the components of the electric and magnetic fields, arranged in a Lorentz covariant way.
where $B_\mu$, $W_\mu^j$, and $G_\mu^\alpha$ are the gauge fields for the hypercharge, weak, and strong interactions, respectively. $Y$, $T_j$ and $S_\alpha$ are the infinitesimal generators for $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$ transformations, respectively. The covariant derivative provides interaction terms between the fermions and the bosons. In principle an infinite number of irrelevant terms can be added. As the operatorial part of the irrelevant terms has mass dimension larger than four, let us call it $k$, these terms are suppressed with $1/\Lambda^k$, where $\Lambda$ is some energy scale and $n = k - 4$. At energies much smaller than $\Lambda$ the effects of these terms are negligible with respect to the others, as they are suppressed by $(E/\Lambda)^n$. As there is no experimental evidence of these terms, i.e. there are only upper limits on the values of the coefficients of these terms, they are not put inside the SM Lagrangian. Such terms would provide evidence of new physics phenomena at a scale $\Lambda$, that manifest in an approximate way at lower energies. This is exactly what happens in the case of the weak interaction at energies much smaller than the mass of the $W$ boson, for instance in the case of the muon decay.

However, since all the terms in the Lagrangian must be gauge invariant, there is an important drawback: no mass term for the gauge bosons and for the fermions is allowed, in contrast with the experimental observations of the $W$ and the $Z$ bosons, and of the fermions. Moreover, another problem arise when thinking about conserved charges: in nature the electromagnetic and colour charge are conserved, and the physical states are organized in multiplets of these charges. Nevertheless the description given so far of the SM seems to predict four conserved charges in the EW sector, namely the hypercharge and the three weak charges related to three $SU(2)_L$ generators, but only a combination of these four, the electromagnetic charge $Q$, is conserved in nature. There are several ways to address these problems using a spontaneous symmetry breaking mechanism. Furthermore there is another constraint, which comes from the experimental observation of the Higgs boson, achieved in 2012 by the ATLAS (Ref. [7]) and the CMS (Ref. [8]) experiments. This scalar and electromagnetic uncharged boson is not present in the theory as described above.

For these reasons, this cannot represent the complete picture: the Higgs mechanism is the last piece of the mosaic.

### 1.2 The origin of the mass: the Higgs mechanism

The mechanism that solves the problems mentioned before is called Higgs mechanism (Ref. [9], [10], and [11])\(^2\), and it is based on the idea that $SU(2)_L \otimes U(1)_Y$ is spontaneously broken. Spontaneous symmetry breaking (better referred as symmetry realized a la Nambu-Goldstone) is a mechanism for which a symmetry of the Lagrangian is not reflected in the ground state, and in the corresponding spectrum that is built starting from it. This is what usually happens at a phase transition. In order to clarify, one can think about a ferromagnet. The Lagrangian that

\(^2\)Actually several scientists published similar works concerning this mechanism at the same time, and therefore, it should be more correct to refer to this mechanism as the Englert–Brout–Higgs–Guralnik–Hagen–Kibble mechanism.
1.2 The origin of the mass: the Higgs mechanism

describes the interaction between the various atoms is invariant under rotations, since there is no
intrinsically preferred direction. Above the Curie temperature, the ferromagnet is invariant under
rotations and the symmetry is realized in the ground state. When the magnet temperature goes
below the Curie temperature, a phase transition occurs, and a non-zero magnetization arises.
The direction of this magnetization is random, but breaks the symmetry of the state under
rotations, although the Lagrangian is not changed. Something similar happens with the Higgs
mechanism. Let us add to the SM Lagrangian the following terms, which do not break any of
the previous symmetries and are renormalizable:

$$L_{\text{Higgs}} = D_\mu \phi^\dagger D^\mu \phi + \psi_i Y_{ij} \psi_j \phi + \text{h.c.} + V(\phi)$$  \hspace{1cm} (1.4)

where $\phi$ is the Higgs field, a complex scalar field, and a doublet under $SU(2)_L$, that has four
real degrees of freedom. The first term is the kinetic term, the second one, with its hermitian
conjugate, is the Yukawa term that characterizes the interactions with the fermions, and the last
term is the Higgs potential

$$V(\phi) = -\frac{1}{2} \mu^2 \phi^\dagger \phi + \frac{1}{4} \lambda (\phi^\dagger \phi)^2$$  \hspace{1cm} (1.5)

with $\mu$ and $\lambda$ real and positive numbers. This functional form for the Higgs potential is unique:
in fact only terms with an operatorial part containing an even number of Higgs fields are allowed,
because the terms with an odd number of Higgs fields are not $SU(2)_L$ invariant, and terms with
more than four Higgs fields are not renormalizable. The key feature of the potential is that it
has a maximum at $\phi = 0$. Consequently, in the vacuum state, the expectation value of $\phi$ is not
zero, but lies somewhere on the circle of minima defined by:

$$\phi^\dagger \phi = -\frac{\mu^2}{\lambda} = v^2$$

The Higgs field acquires a vacuum expectation value (usually referred to as vev). However, this
puts a degeneracy in the ground state, since this condition identifies the minimum up to a phase.
This breaks the global $SU(2)_L \otimes U(1)_Y$ symmetry, as the vacuum state is no longer invariant
under this transformations. There is a combination of the $SU(2)_L \otimes U(1)_Y$ that remains unbroken
though. This is the electromagnetic charge $Q = \frac{1}{2}(Y + T_3)$: in fact, $SU(2)_L \otimes U(1)_Y$ is broken
apart from the subgroup generated by $Q$, that is $U(1)_{\text{em}}$.

The Higgs field can be re-written as:

$$\phi = \frac{1}{\sqrt{2}} e^{\frac{i}{2}(iT_1 \pi_1 + iT_2 \pi_2 + i(T_3 - Y) \pi_3)} \begin{pmatrix} 0 \\ v + h \end{pmatrix}$$

where the $\pi_i$ are the Goldstone bosons arising from the breaking of the symmetry, and $h$ repre-
sents the excitation with respect to the ground state. The idea of this decomposition is sketched
in figure 1.2.

Since the phase of the Higgs field is gauge-dependent, thanks to gauge invariance it is possible
to eliminate this phase. This can be done by redefining the gauge fields in order to absorb the
Figure 1.2: A sketch of the Higgs potential, usually called Mexican hat potential, due to its shape. The minimum of the potential is represented by continuum values, hence there is a continuous infinity of degenerate vacuum states, distinguished by the value of the phase. The underlying symmetry is exact though, and all the vacua are in fact physically equivalent. The choice of one of the vacua breaks the symmetry, and the Higgs field can be rewritten with a physical Higgs boson $h$, that represents the excitation along the "radial" axis, and three $\pi_i$ Goldstone bosons. The $\pi_i$ are mass-less excitations: schematically, they are moving along the continuum minima of the potential, and thus they do not require additional energy as a massive excitations.

Goldstone bosons that thus disappear from the Higgs phase. This gauge choice is called unitary gauge:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}$$

The Goldstone bosons have not disappeared, they are incorporated inside the gauge fields. Computing all the terms in the Lagrangian it is possible to show that from the Higgs potential $V(\phi)$ a mass term for the physical Higgs boson $h$ arises, together with self-interaction terms with three and four $h$'s. From the kinetic term of the Higgs Lagrangian arise both a kinetic term for the Higgs boson $h$, and mass terms for the gauge fields. The three gauge fields related to the broken generators acquire mass, and become the so called $W^+$, $W^-$, and $Z$ bosons, whereas the photon, that is the boson related to the unbroken $U(1)_{em}$ group remains massless. In the unitary gauge, the Goldstone bosons $\pi_i$ become the longitudinal part of the $Z$ and $W$ bosons: indeed, a spin one massless particle has only two degrees of freedom, but after acquiring mass, the longitudinal part appears too. The $Z$ boson and the photon are linear combinations of the $W^3$ and $B$ bosons. The angle that describes this unitary transformation is called weak mixing angle $\theta_W$, and it is related to the coupling constants by:

$$\cos \theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}.$$  

The weak mixing angle describes also the relation between the $Z$ and the $W$ mass:

$$M_W = M_Z \cos \theta_W$$
Interaction terms between the physical Higgs boson \( h \) and the gauge fields also appear from the kinetic term.

From the Yukawa term that is the second one in the Higgs Lagrangian 1.4, it is possible to obtain Dirac mass terms for the fermions and and interactions between them and the physical Higgs boson \( h \). The masses of the fermions are proportional to the vev through the Yukawa couplings. However, since the Yukawa matrix \( Y_{ij} \) is not diagonal for the quarks, the states with defined mass are not the ones interacting with the \( W^+ \) and \( W^- \) bosons. As a result, the transformation matrix between the two basis is a unitary complex 3x3 matrix (called CKM matrix) that can have a complex phase. This complex parameter in the Lagrangian breaks CP, accommodating CP violation in the SM in a natural way.

The experimentally measured value of \( v \) is 246 GeV. This is the only energy scale present in the SM, all the other terms are marginal and thus they do not contain any factor with non-zero mass dimension. Usually one refers to \( v \) as the EW scale.

It is noteworthy that the value of \( v \) was known with high precision several years before the Higgs boson discovery. In fact, the Fermi constant \( G_F \), that is the coupling constant of the weak decays at low energy, is related to \( v \) via \( v = (\sqrt{2}G_F)^{-1/2} \). In the Fermi theory in fact the exchange of a virtual \( W \) is approximated with a four fermions point-like interaction. This approximation is well respected in the low energy weak decays, like the muon decay. From its lifetime it is possible to extract a very precise measurement of the Fermi constant \( G_F = 1.1663787 \times 10^{-5}\text{GeV}^{-2} \).

The Higgs boson discovery allowed instead to set the value of the \( \mu^2 \) of the Higgs potential, that is related to the Higgs boson mass.

### 1.3 Electroweak precision tests

As mentioned in the very beginning, the SM cannot be the end of the story. There are several observations and open questions that requires new physics (NP) beyond the standard model. Moreover there are at least two theoretically problems in the SM as we know it, which prevent to extend it to all the physics scales. One of these is the fact that gravity, as we know it today, would become strongly coupled at the Planck scale, \( M_{\text{Planck}} \approx 10^{19} \text{GeV} \). There are two scenarios: either there are new particles which prevent the amplitudes to become of the order one, or new bound states arise from the strong dynamics of the interaction at the Planck scale. Additionally, even without gravity, at least the coupling constant of the hypercharge would exhibit the Landau pole at very high energy.

NP may be discovered in two complementary ways: direct searches of new phenomena, as the discovery of a new resonance, or by finding an incompatibility between a precise SM prediction and precise measurements. The latter is the result of a precision test: a comparison between the SM prediction and a precise measurement of a certain physical observable. In fact the Standard Model is a predictive theory: once fixed the parameters of the Lagrangian with a finite number of proper measurements, all the physical observables can be computed. Performing precision tests is a test of the consistency of the theory: a disagreement between prediction and measurements
might unveil possible new physics effects that affect the observables through loop corrections.

In this section the importance and the power of the EW precision tests will be illustrated, focusing particularly on two historical examples. In fact the EW sector is one of the most attractive for precision tests: this is because it plays a key role inside the SM, and because both theoretical calculations and experimental measurements are possible with great precision. Performing EW precision tests is a way to both test the consistency of the theory, and possibly unveil quantum effects due to new physics. Indeed, the effect of NP can be parametrized, with the effective field theory approach (EFT), using a set of phenomenological parameters: the value of these parameters can be fitted from measurements, and they can possibly suggest hints of NP phenomena. This is related to the previous reasonings about the SM Lagrangian: in principle all the terms which do not break any symmetry are allowed. The phenomenological parameters are related to the coefficients of the irrelevant operator under consideration, whose values can be predicted with EW precision tests. It is important to keep in mind that precision tests can be sensitive to NP at higher scale than what can be directly proven at the LHC. However, they are always sensitive to $\Lambda/y$, where $\Lambda$ is the new physics scale, and $y$ is the coupling to the EW sector. On the other hand, direct searches explore the scale $\Lambda$, but the results of course depend on the couplings $y$.

The computable quantities are called electroweak precision observables (EWPO) and contain, for instance, the mass of the weak bosons, the weak mixing angles, and the asymmetries in the fermionic decays of the $Z$. Thanks to the structure of the SM, at tree level, all these quantities can be computed starting from only three parameters: the coupling constant of the hypercharge $g_1$, the coupling constant of the weak interaction $g_2$ and the vacuum expectation value $v$. Considering also loop corrections, three more parameters are needed: the strong coupling constant $\alpha_S$, $\mu^2$ of the Higgs potential and the top quark mass $m_t$. In principle the mass of all the fermions are needed for computing loop corrections, however only the top quark contributions are relevant as is much more strongly coupled to the Higgs than the other fermions.

As these parameters are not of easy experimental access, usually a different combination is considered: the electromagnetic constant $\alpha_{em}$, the Fermi constant $G_F$, and the $Z$ boson mass $M_Z$ at tree level; the strong coupling constant, the the top quark mass $m_t$, and the Higgs mass $m_H$ are added for computing loop corrections.

In table 1.1 the values of the six inputs for the EW precision tests are shown, together with their expressions in terms of the parameters of the SM Lagrangian, and of the latest experimental values with their uncertainties as taken from the Particle Data Group, Ref. [2]. The precision on the two quantities $\alpha_{em}$ and $G_F$ comes from the measurement of the magnetic momentum of the electron and the muon lifetime, respectively. The $Z$ mass is known from LEP measurements with a precision of $2 \times 10^{-5}$, the strong coupling constant is known precisely at the energy of the $Z$ mass, and The top quark and Higgs boson masses are known from the Tevatron and LHC measurements.

Since most of the EWPO are computed at the $Z$ mass, it is necessary to use as input the coupling constants at the $Z$ mass: whereas $\alpha_S$ is more precisely measured at the $Z$ pole than at lower
1.3 Electroweak precision tests

Table 1.1: The six input parameters of the EW precision tests, together with their expressions in terms of the parameters of the SM Lagrangian and of the latest experimental values, as taken from [2].

<table>
<thead>
<tr>
<th>quantity</th>
<th>expression in terms of the SM parameters</th>
<th>experimental value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{em}$</td>
<td>$\frac{1}{4\pi \frac{g_1^2}{g_2^2} + \frac{1}{g_2^2}}$</td>
<td>$1/137.035999139(31)$</td>
</tr>
<tr>
<td>$G_F$</td>
<td>$\sqrt{\frac{2}{3}v^2}$</td>
<td>$1.1663787(6) \times 10^{-5}\text{GeV}^{-2}$</td>
</tr>
<tr>
<td>$M_Z$</td>
<td>$\frac{v}{2} \sqrt{\frac{g_1^2}{g_2^2} + \frac{1}{g_2^2}}$</td>
<td>$91.1876 \pm 0.0021$ GeV</td>
</tr>
<tr>
<td>$\alpha_S(M_Z)$</td>
<td>$\frac{g_3^2}{4\pi}$</td>
<td>$0.1179 \pm 0.0012$</td>
</tr>
<tr>
<td>$m_t$</td>
<td>$Y_t v$</td>
<td>$173.34 \pm 0.76$ GeV</td>
</tr>
<tr>
<td>$m_H$</td>
<td>$\mu$</td>
<td>$125.09 \pm 0.24$ GeV</td>
</tr>
</tbody>
</table>

energies, this is not true for $\alpha_{em}$ that must be evolved from $q^2 \approx 0$ to the $Z$ pole. It is noteworthy that the measurement of the muon lifetime, a process which happens at $q^2 \approx 0$, represents a measurement of $v$ which does not show any running as a function of the energy. In computing $\Delta \alpha = \alpha_{em}(M_Z) - \alpha_{em}(0)$, it is difficult to take into account hadronic loop contributions, meaning computing $\Delta \alpha_{had}(M_Z)$. Therefore, a way to extract $\Delta \alpha_{had}(M_Z)$ is to exploit the optical theorem to connect the cross section for the process $e^+e^- \rightarrow$ hadrons as a function of the center-of-mass energy to the amplitude for hadronic loops, as shown in Ref. [12]. This procedure results in $\Delta \alpha_{had}(M_Z) = 0.02750 \pm 0.00033$, that is much less precise than the starting value of $\alpha_{em}(0)$. It is important to keep in mind that, even if $\Delta \alpha_{had}(M_Z)$ is needed to compute EWPO at the $Z$ pole, it is not one of the six fundamental inputs of the EW precision tests. In fact, it takes into account loop contributions which do not affect directly the EWPO, but affect the running of the $\alpha_{em}$ from different $q^2$. If a more precise measurement of $\alpha_{em}$ was performed at the $Z$ pole, there would not be any need for $\Delta \alpha_{had}(M_Z)$.

Starting from these six parameters, the analytic expression for the EWPO and their measurements, it is possible to:

- Fix the values of the input parameters and predict all the EWPO: the theory is subsequently tested by comparing the predicted values with the measurements.

- Fit the values of some of the inputs and EWPO starting from the analytic expressions and the experimentally measured values of these quantities, which are added in the likelihood of the fit. This is what it is usually called the EW fit: comparing the fitted values with the measurements test particularly the consistency of the theory. The pulls of the EWPO added in the Likelihood and the $\chi^2$ of the fit are other parameters to test the consistency of the fit result with the initial hypothesis: inconsistency in the fit means inconsistency of the theory, which can be a hint of NP.

- Generalize the SM fit considering different sets of parameters which account for NP contri-
butions in several extensions of the SM. The NP phenomena introduced by extending the SM with NP at higher scale can be parametrized with EFT and irrelevant operators, as previously discussed. The coefficients of these operators can be related to the parameters of the theory which describe NP and their values can be extracted from the fit.

As an example, in the left plot of figure 1.3 it is reported the comparison between the measurement and the prediction of the top quark mass, as a function of the year. Even before the discovery, the EW fit was able to constrain the mass of the top quark around its value, giving a hint of the direction where to search for discovering the top quark.

![Figure 1.3: The two plots show a comparison between the prediction and the experimental measured values of the top quark mass (left) and Higgs boson mass (right) as a function of the year. Even before their discovery, the EW fit was able to predict their masses with good precision and accuracy. Both plots taken from [13], where more information about the EW fit can be found.](image)

Another interesting example is represented by the Higgs mass $m_H$. The right plot in figure 1.3 shows a comparison between the EW fit prediction and the experimental measurement of the Higgs boson mass as a function of the year. Since 1995, around twenty years before its discovery, the Higgs boson mass was predicted with uncertainty of roughly 50-100 GeV starting from the Tevatron and LEP measurements and direct searches. Furthermore, in 2008, before the LHC, the EW fit prediction of the Higgs mass was (Ref. [14]):

$$m_H^{\text{theo}} = 116.4^{+18.3}_{-13.3} \text{ GeV}$$

The precision reached by the EW fit was able to constrain the mass of the Higgs boson in the range where it was eventually found, showing a good consistency of the SM.

1.4 Motivation for a precise $W$ mass measurement

Among the electroweak observables, one of the most important is the $W$ mass. The tree level expression of the $W$ mass is the following:

$$M_W = \sqrt{\frac{\pi \alpha_{em}}{G_F \sqrt{2} \sin \theta_W \sqrt{1 - \Delta r}}}$$
1.4 Motivation for a precise $W$ mass measurement

where it has been expressed in terms of $\alpha_{em}$, $G_F$ and $\sin \theta_W$ for simplicity, but it can be rewritten also in terms of the $Z$ mass (that is in the chosen basis of the input variables for the computation of the EWPO observables). $\Delta r$ represents all radiative corrections: in the SM, the mass of the $W$ boson takes relevant loop corrections from the top quark, that grows quadratically with the top mass, and from the Higgs boson, that grows logarithmically with the Higgs mass. Moreover, in most extensions of the SM it is affected by new physics contributions in an important way.

Furthermore, in this historical moment, the $W$ mass is one of the most interesting EWPO since the experimental precision has not yet reached the theoretical one. Specifically, the world average of the $W$ mass measurements is [12]:

$$M_{W}^{exp} = 80.385 \pm 0.015 \text{ GeV}$$

whereas the prediction of the SM (Ref. [1]) is, as explained as the first method to predict EWPO in the previous section:

$$M_{W}^{theo} = 80.3618 \pm 0.0080 \text{ GeV}$$

where the break down of the uncertainties is reported in table 1.2:

<table>
<thead>
<tr>
<th>Prediction</th>
<th>$\alpha_S$</th>
<th>$\Delta \alpha_{had}$</th>
<th>$M_Z$</th>
<th>$m_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_W$ [GeV]</td>
<td>80.3618 ± 0.0080</td>
<td>± 0.0008</td>
<td>± 0.0060</td>
<td>± 0.0026</td>
</tr>
</tbody>
</table>

The biggest uncertainties come from $\Delta \alpha_{had}$ and $m_t$, whereas the uncertainty on the Higgs mass plays a negligible role when propagated through the logarithmic correction to $M_W$.

Therefore it is of primary importance to lower the experimental uncertainty on the $W$ mass measurement, in order to fully exploit the power of the prediction of the theory.

In table 1.3 both the posterior and prediction of the EW fit for $M_W$ are shown. These two quantities are obtained with the second method of predicting EWPO, as explained in the previous section. The posterior is obtained combining all the experimental measurements in the likelihood of the fit. The prediction is instead the posterior in case that all the experimental measurements are added in the likelihood but $M_W$. The precision reached is of 6.1 and 6.6 MeV for the two cases, respectively. In addition, there is a pull of 1.5$\sigma$ between the measurement and the fitted value. This pull is not yet discrepancy, hence a hint of NP, but it is interesting to lower the precision in order to reduce or possibly enlarge this discrepancy.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Posterior</th>
<th>Prediction</th>
<th>Pull</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_W$ [GeV]</td>
<td>80.385 ± 0.015</td>
<td>80.3644 ± 0.0061</td>
<td>80.3604 ± 0.0066</td>
</tr>
</tbody>
</table>

It is possible to visualize the same result with the 2D plots shown in figure 1.4. In this case
the results of several EW fits, performed by removing subsequently from the Likelihood the measurements of $M_W$ and various quantities related in the fit to the $W$ mass, like $\sin^2 \theta_{\text{eff}}^{\text{lept}}$, $m_H$, $m_t$ and $\Gamma_W$, are shown together with their 68% CL. The effective leptonic mixing angle $\sin^2 \theta_{\text{eff}}^{\text{lept}}$ is determined from the ratio of the dressed vector and axial vector couplings $g_V$, $A$ of the $Z$ boson to the leptons through:

$$\sin^2 \theta_{\text{eff}}^{\text{lept}} = \frac{1}{4} (1 - \text{Re} \frac{g_V}{g_A})$$

and its value is related to the $\sin^2 \theta_W$ via $\sin^2 \theta_{\text{eff}}^{\text{lept}} = k \sin^2 \theta_W$, where $k$ is unity at tree level and incorporates loop contributions.

The left pane shows the result for $M_W$ versus $\sin^2 \theta_{\text{eff}}^{\text{lept}}$, whereas in the right pane the result for $M_W$ versus $m_t$ is plotted. It is possible to see that the theoretical precision is better than the experimental one and that there is a pull of about $1\sigma$ between the measured value and the predicted one.

### 1.5 Conclusions

In this chapter a general overview of the SM has been given, focusing mainly on the EW sector and the spontaneous symmetry breaking mechanism.

The predictiveness of the this theory has been illustrated, in the context of the EW precision tests. Eventually the results of the latest EW precision tests that predict the $W$ mass with high precision, put a strong motivation for a more precise $W$ mass measurement, in order to fully exploit the precision of the theoretical prediction.
Chapter 2

The CMS experiment at the LHC

This chapter contains an overview of the experimental facilities that have been used to collect the data used for the analysis. Since they are among the most complex experiments ever built, and required the efforts of thousands people over a long period, the discussion of the experimental apparatus might be extremely long. Therefore, instead of going in the details of the experiment, a brief overview is presented, focusing on the most important ingredients for performing precision measurements. More details about the LHC and the CMS detector can be found in Ref. [15] and Ref. [16], respectively.

2.1 The Large Hadron Collider

The Large Hadron Collider is a particle accelerator designed to accelerate protons and nuclei to a center of mass energy of $\sqrt{s} = 14$ TeV and $\sqrt{s} = 5.52$ TeV per nucleon, respectively. It is hosted in the LEP tunnel, a circular cavern 26.7 km long in the CERN-Geneva area. The center of mass energy reached in the 2011 proton run was of 7 TeV, in 2012 it was raised to 8 TeV and it eventually reached 13 TeV, which is very close to its nominal value of 14 TeV, in 2015.

Two transfer tunnels link the LHC to the CERN accelerator complex that acts as an injector. The protons start from the LINAC2 then pass through the Booster (PBS), followed by the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) that injects them in the LHC. More details on the CERN accelerator complex and on its features can be found in [15] and a schematic representation is shown in the left plot of figure 2.1.

The LHC is composed of two rings with four crossing points, flanked by long straight sections, as shown in the right plot of figure 2.1. Each straight section has a length of 528 m and can serve as an experimental or utility insertion. The four crossing points are occupied by the experiments. The detectors installed at the interaction points are, reported in the numerical order of the LHC scheme, ATLAS, LHCf (that shares the ATLAS cavern), ALICE, CMS, TOTEM (that shares the CMS cavern) and LHCb.

The beam is not continuous but organized in bunches. They are usually 5 cm long, with a
transverse section that is about 10x100 µm at the interaction points, and they are interspersed 25 ns, corresponding to roughly 7.5 m at the speed of light. First of all, this makes possible to accelerate them in radio-frequency cavities: if the bunch frequency matches the oscillation frequency of the electric fields inside the cavities, the bunches always feel an electric field in their direction of motion and are always accelerated. Secondly, the collisions occur only during "bunch crossings". A bunch crossing usually lasts for about 200 ps. This has the advantage of reading the detector just over a small time around this instant, classifying the collisions in events. Indeed what is called an event is a complete read-out of the detector. However, because of the transit time of the particles through the detector and because of the integration time of sensors electronics used for the read-out, also some signals of particles produced in few previous and subsequent bunch crossings might be recorded.

One of the most important feature of a collider is the luminosity, defined as:

\[ \mathcal{L} = \frac{R}{\sigma} \]

where \( R \) is the rate of events of a given process and \( \sigma \) is the cross section for that given process. \( \mathcal{L} \) is the instantaneous luminosity, usually expressed in cm\(^{-2}\)s\(^{-1}\). For circular machine with approximately Gaussian bunches with transverse dimensions \( \sigma_x, \sigma_y \), it can be written as

\[ \mathcal{L} = f \frac{N_1 N_2}{4\pi \sigma_x \sigma_y} \]

where we assumed two bunches containing \( n_1 \) and \( n_2 \) particles colliding with frequency \( f \). For the LHC \( n_1 = n_2 \simeq 10^{11} \), \( f = 40 \) MHz, \( \sigma_x \simeq 100\mu m \), and \( \sigma_y \simeq 10\mu m \): these features leads to \( \mathcal{L} \simeq 3 \times 10^{33} \) cm\(^{-2}\)s\(^{-1}\).

The LHC has the possibility to deliver large instantaneous luminosity: this is important both to search for rare production and decays, like the Higgs production, and to study with high precision
2.1 The Large Hadron Collider

standard model processes, such as the $W$ and the $Z$ bosons, since large rates of interesting events are produced. Specifically, in 2012, the peak instantaneous luminosity reached $7.7 \cdot 10^{33} \text{cm}^{-2}\text{s}^{-1}$, whereas, during 2016, it exceeded the design luminosity ($10^{34} \text{cm}^{-2}\text{s}^{-1}$), reaching the new record of $1.4 \cdot 10^{34} \text{cm}^{-2}\text{s}^{-1}$.

Since what matters for physics is the number of events and not the rate, it is usually useful to know the integrated luminosity, defined as:

$$L_{\text{integrated}} = \int dt L_{\text{instantaneous}}(t)$$

where the integral runs over a certain period of time. For instance, in 2016, the LHC delivered an integrated luminosity of $40 \text{ fb}^{-1}$ instead of $25 \text{ fb}^{-1}$ expected, because of the increase in the instantaneous luminosity. Given a cross section of 10 nb for the process $pp \rightarrow W \rightarrow \mu \nu$, the rate during 2016 with $L = 10^{34} \text{cm}^{-2}\text{s}^{-1}$, was roughly 100 Hz, and 400 millions events were produced.

The evolution of the integrated luminosity for the various LHC runs is shown in figure 2.2.

![Figure 2.2: Delivered luminosity by the LHC as a function of the day of the year, for the various LHC runs. As it is possible to see, during 2016 the delivered luminosity exceeded 40 fb$^{-1}$, by far larger than the expectation. This was possible thanks to the achievement and following overtaking of the instantaneous design luminosity.](image)

A huge amount of instantaneous luminosity has also an important collateral effect: the increase of the pile-up collisions (PU). Indeed, the luminosity is so high that the average number of proton-proton interactions in each event is larger than one. This is to say, multiple proton-proton interactions take place per bunch crossing, enlarging the number of particles measured in each event and creating ambiguity in separating them in the various collisions. The average
number of collisions during run I was 8(21) at 7(8) TeV, whereas during run II, at 13 TeV, it reached 14(27) during 2015(2016).

2.2 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) detector, is located at the 5th access point of the LHC tunnel in the vicinity of the town of Cessy, in France. It was designed to withstand the LHC running conditions and at the same time to fulfill a rich physics program detailed in Ref. [17]. The main features of the detector, needed to operate in the LHC environment, are: radiation hardness, good time resolution, and capability to recognize interesting events in very short amount of time. These requirements are imposed by the large proton-proton total cross section, the high beam intensity and the high frequency of the collisions. A further necessity is an high granularity, both to reconstruct single particles with high efficiency, in view of the physics goals, and to be able to operate with a high number of PU interactions.

CMS is made of several sub-detectors: by combining the information collected by all of them it is possible to measure with high precision the momentum and the energy of the particles produced in the event, while identifying them thanks to the features of their interactions in the various sub-detectors. An illustration of the various CMS sub-detectors is presented in figure 2.3.

Figure 2.3: An illustration of the various sub-detectors inside CMS.
2.2 The Compact Muon Solenoid experiment

Specifically, starting from the interaction point, located at the very center of the detector, the main sub-detectors are:

- **Silicon tracker**: a silicon tracker, made of fourteen layers of pixels and micro-strips detectors, allows a very precise tracking of charged particles.

- **ECAL**: the electromagnetic calorimeter is a homogeneous calorimeter made of lead tungstate crystals, allowing a very precise measurement of the energy of the electromagnetic showers produced by electrons and photons.

- **HCAL**: the hadron calorimeter is a sampling calorimeter made of brass and scintillator layers, allowing for an energy measurement of the showers produced by hadrons.

- **Superconductive magnet**: as the name suggests, one of the main features of the CMS experiment is its superconductive magnet. It produces a magnetic field of 3.8 T parallel to the beam axis in the core of the detector. In combination with the silicon tracker, it allows a very precise measurement of the transverse momentum of charged particles.

- **Muon system**: the superconductive magnet is surrounded by the steel return yoke of the magnet. Inside the yoke, various gaseous detectors are situated in order to perform precise measurements of muons, which are the only visible particles able to penetrate this area of the detector. These are drift chambers (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC).

**Notation of kinematic quantities**

For later convenience we report the CMS coordinate system. It is a right-handed coordinate system, with the origin at the nominal interaction point, the x-axis pointing to the centre of the LHC, the y-axis pointing up, perpendicular to the LHC plane, and the z-axis along the anticlockwise beam direction. The polar angle, \( \theta \), is measured from the positive z-axis and the azimuthal angle \( \phi \), is measured from the x axis.

It is usually more convenient to consider a different set of variables instead of energy and angle or Cartesian coordinates. In fact, the invariant phase space can be written as:

\[
d^4p \delta(E^2 - p^2 - m^2) = p_T dp_T d\phi dY
\]

Where the **transverse momentum** \( p_T \) is the component of the momentum perpendicular to the beam axis, and the **rapidity** (along the beam axis) is defined as:

\[
Y = \frac{1}{2} \log \frac{E + p_z}{E - p_z}
\]

where \( p_z \) is the component of the momentum along the beam axis. In case of massless particles (E≈m), that goes produced at large angles with respect to \( 1/\gamma \) (\(|\sin(\theta/2)| >> 1/(2\gamma))\), the rapidity coincides with the pseudorapidity defined as:

\[
\eta = -\log(\tan(\theta/2))
\]
Chapter 2. The CMS experiment at the LHC

and it is a biunivocal correspondence with the polar angle $\theta$. This choice is due to the fact that $p_T$ is invariant under boosts along the beam axis, whereas $Y$ is additive for the same boosts. Another important aspects is that the production of hadrons in soft QCD collisions is almost flat in the $\eta$ variable.

Physics objects and identification  The identification of the particles is a crucial point. Indeed the detector does not provide particles as output, but provides hits in the tracking system and energy deposits in the calorimeters. Hence, physics objects are reconstructed from the output of the detector (see the particle flow section in this chapter for more information). For example, a quark or a gluon in the final state appears in the detector as a narrow cone of hadrons. It happens because of QCD confinement: coloured particles cannot exist free. This is what it is usually called jet. From an experimental point of view, a jet is reconstructed using a certain algorithm for clustering the reconstructed particles in the final state. A muon instead is obtained by matching hits in the tracker and in the muon system. Neutrinos, as well as other possible weakly interacting particles not yet discovered, are revealed by absence of momentum, called missing transverse energy (MET). In fact the initial state of the collision is well known, with zero total momentum, which is conserved in the final state. However, a measurement of the longitudinal component is not possible because of the huge amount of momentum lost in the forward direction, outside of the detector acceptance. A measurement of the transverse momentum is instead possible, because the particles that are lost usually do not carry significant $p_T$. The $E_T^{\text{miss}}$ is thus computed as the negative vectorial sum of the transverse momenta of the revealed particles.

This process can be done in different ways since the definitions are not unique. Some "fast" definitions can be used at the trigger level (see the next paragraph for more details), and some more complex can be used offline for obtaining better performance.

Isolation  The isolation is another important concept. For distinguishing between objects that are produced inside jets and objects that comes alone in the final state (from the decay of a resonance for instance), it is useful to look at the energy carried by other particles surrounding the particle of interest. The isolation is thus defined as the sum of the energy (or the $p_T$) of the particles in a given cone ($\sqrt{\Delta \phi^2 + \Delta \eta^2}$) around the particle of interest. For instance, a muon coming from the $W$ decay is expected to be isolated, whereas a muon produced in the decay of a B-meson will be more likely surrounded by hadrons and photons. This variable is important both for triggering and for selecting the events for the analysis.

Trigger  A fundamental feature of the experiments at hadron collider is the trigger: that is the system that rapidly decides which events to record since only a small fraction of the total produced events can be stored. All the cross sections related to interesting processes, such as EW, top quark production, Higgs boson, possible exotic new physics, are extremely small in comparison to the total proton-proton cross section. Specifically, the total pp cross sections is
2.2 The Compact Muon Solenoid experiment

about 100 mb, whereas for Z or W production is around 10-100 nb, and it decreases to 10 pb for Higgs production. Therefore a huge instantaneous luminosity is needed in order to produce significantly these rare processes. For instance, only $2 \cdot 10^5$ Higgs events were produced in the 2012 run; if we search for it in the diphoton decay channel (branching fraction $\simeq 0.2\%$), only 400 events are expected. For new physics beyond the standard model, like supersymmetric particles, the predicted cross sections are even smaller. For example, the predicted cross section for producing squarks, with mass of 600 GeV, is around 0.1 pb: this reflects to roughly 2000 expected events in the whole 2012 run. However, it is not possible to store all the events that are produced: the rate of bunch crossings is 40 MHz, one every 25 ns, whereas roughly only 1000 events can be stored every second, meaning one every 40000 bunch crossings. It is thus extremely important to recognize the interesting events online and with high efficiency: this goal is achieved in CMS with two trigger steps.

The first one is called Level 1 (L1) trigger. It is the first step of the trigger. It is based on FPGAs, and exploits the information coming from the calorimeters and from the muons system (it is impossible to read out and store the tracker at 40 MHz). This trigger analyzes the events one after the other, taking in buffers the whole information of the events, while only a minimal part of the read-out is used at L1. This trigger reduces the rate from the beam crossing frequency to 100 kHz. The events selected by this trigger are classified based on the feature they triggered: for instance muons, hadron calorimeter energy, ECAL deposits or missing transverse momentum.

The second trigger level is called high level trigger (HLT). It is based on a CPU farm that processes all the events selected by the L1 trigger in a parallel way. Each event that passes the L1 trigger is analyzed by a computer, independently on the other events, which reconstructs the features of interest. For instance, the identification of particles, like electrons and photons, the isolation or the invariant mass of two particles can be computed at this level using the full event information. The trigger is organized in a menu with several trigger paths, each of them consisting in a series of steps that, after computing certain quantities, decide if discarding or not the event. The events that reach the end of at least one line can be stored, together with the information of the lines they triggered.

2.2.1 The superconductive magnet

As previously mentioned, the main feature of the CMS detector is a superconductive solenoid magnet, capable to produce an uniform magnetic field of 3.8 T over a length of 12.5 m and on a radius of 3.15 m. Inside it the tracker and both the electromagnetic and hadron calorimeters are accommodated.

2.2.2 The silicon tracker

The CMS silicon tracker is a cylinder-shaped detector, with length 5.6m and radius 1.2m, measuring charged particles up to $|\eta| = 2.5$. It is divided into two parts: a barrel and two endcap parts. The barrel is composed of three layers of pixels and ten layers of micro-strips, whereas
Chapter 2. The CMS experiment at the LHC

each of the two endcaps contains two layers of pixels and twelve layers of micro-strips \(^1\). A schematic representation of the CMS tracker is shown in figure 2.4. The entire tracker consists of 16588 sensor modules, that are segmented into 66 million 150x100 µm pixels and 9.6 million 80-to-180 µm-wide strips. The fine granularity provided by the CMS tracker offers a great separation between close tracks, particularly in the pixels region, where, near the vertices, the flux is extremely high. In spite of the high flux, this fine granularity allows to keep the occupancy between 0.1 % and a few %. More information about the CMS tracker and its performance can be found in [18].

![Figure 2.4: An illustration of the CMS tracking system. The acronyms TIB, TID, TOB, and TEC stand for "tracker inner barrel", "tracker inner disks", "tracker outer barrel", and "tracker endcaps", respectively.](image)

The material budget, that is the amount of material expressed in radiation lengths or nuclear interaction lengths, is shown in the left pane of figure 2.5, as a function of the pseudorapidity \(\eta\). This is closely related to the probability, for a photon or an electron, of converting or emitting bremsstrahlung radiation. For a hadron, the material budget expressed in number of nuclear interaction lengths tells the probability of experiencing a nuclear interaction. The material budget is also related to the multiple scattering of the charged particles, that affects the resolution negatively both in the measurement of the transverse momentum and impact parameter, and in the reconstruction of the vertices.

One aspect that is particularly important in this thesis is the capabilities of the CMS tracking system of reconstructing tracks with small transverse momentum. As it is possible to see in the right pane of figure 2.5, after the full CMS reconstruction, the tracking efficiency for charged hadrons as a function of \(p_T\) increases linearly between 250 and 900 MeV, where it stabilizes at its constant value of 0.9 up to 20 GeV. In this case the main limit is due to the non-zero probability, for a charged hadrons, of experiencing a nuclear interaction in the tracker. Indeed, for muons, \(^1\)actually the pixels detector changed at the beginning of 2017, and this description refers to the previous one.

---

\(^1\)actually the pixels detector changed at the beginning of 2017, and this description refers to the previous one.
2.2 The Compact Muon Solenoid experiment

Figure 2.5: Left pane: total thickness of the CMS tracking system expressed in radiation lengths, as a function of the pseudorapidity (the material budget is almost flat in the azimuthal φ coordinate). The meaning of the acronyms is the same as for figure 2.4. This figure is taken from [18].

Right pane: tracking efficiency for charged hadrons as a function of the transverse momentum. The full offline reconstruction (red line) increases significantly the efficiency with respect to the first iteration (black line). This figure is taken from [19], where it is possible to find more details about this methodology.

the tracking efficiency stabilizes at the value of 0.99, very close to unity.

Another important feature of the CMS tracking system, is the excellent resolution in the measurement of the transverse momentum and of the impact parameter of the tracks. In the left pane of figure 2.6 the relative resolution on the transverse momentum measurement is shown, as a function of the transverse momentum and for single, isolated muons. In the region of interest for the W mass analysis, muon with transverse momentum around 40 GeV, the resolution at 68% (solid symbols) is is below 2% in most of the angular region. In the right pane of figure 2.6 the resolution on the parameter $z_0$, that is the impact parameter on the z-axis, is shown for isolated pions as a function of the pseudorapidity. With pions with transverse momentum relatively high, as in high energetic jets, the resolution is of the order of few tens of $\mu$m. This makes possible to reach excellent performance in the tagging of the b-jets through the analysis of the impact parameters of the tracks. For the $M_W$ analysis, where soft hadrons are produced, a good $z_0$ resolution is important for the correction vertex association. Considering a luminous region of 5 cm with vertices homogeneously distributed, the $z_0$ resolution between 100 and 1000 $\mu$m is smaller than the average distance between two vertices up to pileup between 500 and 50.

2.2.3 The electromagnetic calorimeter

The CMS electromagnetic calorimeter is composed of PbWO$_4$ crystals arranged in a barrel section (EB) and in two endcap sections (EE). The former covers the azimuthal coordinate up to $|\eta| < 1.479$, whereas the latter extends from there to $|\eta| < 3$. The crystals have a truncated
Chapter 2. The CMS experiment at the LHC

Figure 2.6: Left pane: the relative transverse momentum resolution, as a function of the transverse momentum, is shown for single, isolated muons. The barrel, transition, and endcap regions, defined by \( \eta \) intervals of 0–0.9, 0.9–1.4 and 1.4–2.5, respectively. The solid (open) symbols correspond to the 68% (90%) intervals.

Right pane: the resolution on the measurement of the track parameter \( z_0 \), as a function of the pseudorapidity, is shown for single, isolated pions. The solid (open) symbols correspond to the 68% (90%) intervals.

Both figures are taken from [18].

Pyramidal shape. They are oriented towards the interaction point, forming a quasi-projective geometry, with a slight tilt of 3 degrees with respect to the line joining the vertex with the center of their square face. The crystals, although transparent to visible light, are extremely dense. With a radiation length of 0.89 cm and a length of 23(22) cm in the barrel/endcap, they offer roughly 25 radiation lengths to electromagnetically interacting particles. This large number is crucial to completely contain the electromagnetic showers. As the crystals contain also one nuclear interaction length, about two thirds of the hadrons start showering inside the ECAL. The crystal transverse size matches the Molière radius of 2.2 cm. This fine granularity of the detector allows the separation of hadron and photon energy deposits that are not closer than 5 cm, allowing a single particle reconstruction even inside jets.

The resolution of the calorimeter, measured with electron beams, is parametrized as a function of the energy:

\[
\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\%
\]

The first term in the resolution formula is the so called stochastic term, and it is related to the intrinsic fluctuations of the shower (that is a stochastic process). The behavior \( \simeq 1/\sqrt{E} \) reflects the stochastic fluctuations in a Poisson process that go like \( \Delta N/N \simeq 1/\sqrt{N} \). As a property of homogeneous calorimeters, the stochastic term is very small. This allows an excellent energy resolution in the typical range of the photon energy inside jets, normally between 1 and 50 GeV.
Furthermore, for the purpose of the $W$ mass measurement, this ensures a good energy resolution for the electrons coming from the vector bosons decay, which is typically between 20 and 60 GeV. The response of the crystals depends strongly on the transparency of the crystals, that changes during the run because of the radiation. This is continuously monitored with a laser system. Each of the crystals is equipped with a laser that shoots a pulse and measures the response roughly every 40 minutes. In addition, for correcting the response (as a function of time) and calibrating the calorimeter, the symmetry of the energy flux in the $\phi$ direction and the photon pairs coming from the $\pi_0$ decay are exploited. This calibration and equalization of the response of the various crystals is fundamental in order to exploit the excellent energy resolution of the ECAL. A different response between the various crystals would result in an effective resolution term proportional to the energy, that would significantly degrades the resolution.

2.2.4 The Hadron Calorimeter

The HCAL is a hermetic sampling calorimeter made of several layers of brass and plastic scintillator tiles. It is divided in a barrel, covering the range $|\eta| < 1.3$ and two endcaps $1.3 < |\eta| < 3$. This provides about six interaction lengths in the central region and increases to over ten interaction lengths at larger pseudorapidities. A forward part of the hadron calorimeter (HF) is situated at $\pm 11$ m from the interaction point, covering pseudorapidities up to $|\eta| = 5$.

The important features of the HCAL is the containment of the showers and the hermeticity of the detector, important for measuring the missing transverse momentum, taking into account that the HCAL is almost totally placed inside the magnet, putting additional geometrical constraints. In fact, a tail catcher (HO), installed outside the solenoid, is used to measure the tails of the hadronic showers. Taking into account the ECAL and the solenoid, the total depth of the calorimeter system is extended to a minimum of twelve(ten) interaction lengths in the barrel/endcaps.

The HCAL resolution, measured with a pion test beam, results to be:

$$\frac{\sigma}{E} = \frac{110\%}{\sqrt{E/\text{GeV}}} \pm 9\%$$

although the resolution is not the best feature of the HCAL, thanks to the possibility of using the particle flow algorithm (discussed in the detail in the following section) to measure the charged tracks and the photons inside jets, the poor resolution of the HCAL affects only the measurement of the neutral hadrons, that are responsible for around 10% of the typical jet energy.

This detector, given its fast read-out is also very important for L1 trigger, where it is possible to set up triggers on the energy of a jet as well as on the missing transverse momentum.
2.2.5 The muon system

Since the muons are the only visible particle traveling the entire detector without experiencing any strong interaction, the muon detectors are installed in the yoke for the magnetic flux return. The yoke consists of three layers of steel, in which four muon detector planes are installed. Drift tube (DT) chambers and cathode strip chambers (CSC) detect muons in the regions $|\eta| < 1.2$ and $0.9 < |\eta| < 2.4$ respectively. In addition, a system of resistive plate chambers (RPC), covers the region $|\eta| < 1.6$. The RPCs operate in avalanche mode. Although the poor spatial resolution, since they are very fast, they turn out to be extremely useful for fast triggering.

The momentum resolution takes advantages of the combination of the tracker and the muon system measurements. Specifically, at small values of the transverse momentum, the tracker resolution dominates, and the muon system is mainly used for triggering and for identifying muons; whereas, at high values of transverse momentum the measurement performed with the muon system becomes important. This transition happens around 200 GeV.

Furthermore, given the big quantity of material present in the detector, a muon usually loses 5-6 GeV traveling through the muon system. Therefore, low energy muons are usually measured only in the tracker, and can be easily misidentified as charged hadrons. This is not important for the energy regime of the W mass though, where the muons coming from the vector bosons decay, have typical transverse momentum larger than 20 GeV. Hence, they are usually well identified, and well measured in the tracking system. More information about the performance of the CMS muon detector can be found in [20].

2.2.6 The particle flow reconstruction algorithm

In the modern concept, general-purpose detectors are generally organized in a cylindrical way around the interaction point, and the various sub-detectors are arranged in such a way to measure the various particles produced in an event. Specifically, starting from the interaction point, the tracking system is the first element which is crossed by the particles. The trajectories of charged particles are reconstructed from hits in the tracking layers, and subsequently grouped together in order to reconstruct the vertices. Then, the electromagnetically interacting particles are reconstructed in the electromagnetic calorimeter as energy clusters. Hadrons can start a hadronic shower either in the ECAL or in the HCAL: a hadron can therefore leave its energy either entirely in the ECAL, or entirely in the HCAL, or in both ECAL and HCAL. Eventually, muons, after traveling the whole detector almost untouched, leave additional hits in the muon system, whereas the neutrinos escape the detectors with no interaction. This is schematically sketched in figure 2.7.

Traditionally at hadron colliders, each physics object is reconstructed in one of the sub-detectors. Jets, made of hadrons and photons, are inclusively measured in the calorimeters, without attempting to separating the jet content. The same way applies to the missing transverse energy ($E_T^{\text{miss}}$) reconstruction. The electrons and photons reconstruction mainly relies on the ECAL. The tracker is mainly exploited for tagging the jets produced by b quarks hadronization or
2.2 The Compact Muon Solenoid experiment

Figure 2.7: An illustration of the interactions of different particles with the various sub-detectors inside CMS.

by hadronic $\tau$ decays. The muon identification is performed with the muon system, and its measurement relies also on the tracker.

The particle flow (PF) algorithm aims to improve the description of the event by combining together the basic elements: tracks and clusters. This combination is used to identify and then reconstruct each particle individually, as opposed to the traditional way. First of all, individual particles in the event are identified, thanks to their expected specific interactions in the successive subdetectors. Secondly, the corresponding measurements of the properties of a particle in the various subdetectors are combined together. Specifically, neutrinos do not interact in the subdetectors; muons are identified as tracks, followed by almost no deposit in the calorimeters and possibly a track in the muon system; electrons are identified as tracks followed by an ECAL deposit compatible with the momentum of the track and no HCAL cluster; photons as clusters in the ECAL with no track and no deposit in the HCAL; charged hadrons as tracks plus ECAL, or HCAL or two compatible deposits in both ECAL and HCAL; neutral hadrons as clusters in the ECAL, in the HCAL, or in both the calorimeters. Thanks to this algorithm the charged hadron energy resolution can exploit the superior resolution of the tracker for the charged hadrons, or the calorimeter resolution in case of badly reconstructed tracks. The clusters related to neutral particles can be better calibrated under the hypothesis of photon or neutral hadron. The energy resolution of the electrons is dominated by the tracker at low energy and by the calorimeter at higher energy. The muon resolution benefits of the tracker measurement at small transverse momentum, and exploits the bigger lever arm of the muon system at higher momentum.

The PF algorithm gives as output a list of reconstructed particles, usually called particle flow candidates. Physics objects are then built from this list of particles, improving considerably the resolution on the measurement of their properties. Jets are obtained by clustering together
hadrons and photons, and in an analogous way the missing transverse momentum is reconstructed. The energy resolution of jets and missing transverse momentum is thus improved up to a factor two, and the angular one up to a factor four. Additionally, the precise knowledge of the particle content allows pileup mitigation, boosted jet substructure analysis, and tagging of heavy flavours. Then hadronic $\tau$ decays can be reconstructed by clustering together hadrons and photons. Isolated photons and leptons are also reconstructed from the list of PF candidates, by exploiting the full event description to compute the isolation.

The most important features of the CMS detector that allows this type of reconstructions are the following. Firstly, the granularity of the detector, both the tracker and the ECAL, allows a fine separation between clusters produced by close particles. Secondly, the presence of the calorimeters inside of the magnet, reduces significantly the amount of material in front of the calorimeters. This allows tracks reconstructed in the tracker and clusters in the calorimeters to be linked in a better way. Furthermore, a PF algorithm requires to understand in detail the detector qualities, to master the reconstruction of the tracks and clusters, and to link them together in a clever manner.

The first development of a particle flow algorithm was in ALEPH (one of the four detectors at LEP): a short description of the method can be found in [21]. A complete description of the particle flow reconstruction algorithm and an illustration of its performance with the CMS detector during run I can be found in this legacy paper [19]. It is noteworthy that, nowadays, the PF reconstruction algorithm is developed for each of the future high energy physics experiments, particularly for those studying $e^+e^-$ collisions.

### 2.3 Conclusions

In this chapter a brief overview of the experimental settings that has been used for the $W$ mass measurement has been given, focusing on the aspects that make CMS a suitable detector for this precision measurement. Specifically, the excellent lepton identification and momentum resolution ans the fine granularity of the detector, that allows single particle identification and reconstruction, are two of the key ingredients for this measurement. This concept will be better explained in the next few chapters.
Chapter 3

Measuring the $W$ mass

In this chapter an overview of the $W$ mass measurement will be presented, starting from its historical evolution during the last forty years, to the description of the measurement at the hadron colliders, and the main challenges at the LHC for reaching the most precise measurement.

3.1 A historical overview

The $W$ boson was first predict in 1967 by Sheldon Glashow, Steven Weinberg, and Abdus Salam in the first works on the modern EW theory (Ref. [22] and [23]), leading to the Noble Prize for the three scientists in 1979. Fifteen years later, in 1982, its discovery at the experiments UA1 and UA2 (Ref. [24] and [25]) at the Super Proton Syncrotron at CERN, led to the Noble Prize for Carlo Rubbia and Simon van der Meer. Since then, its properties, and particularly its mass, have been studied continuously, both at lepton and hadron colliders. The first measurements of the $W$ mass were performed by the UA1 and UA2 collaboration, giving a result of $81 \pm 5$ GeV, and $80^{+10}_{-6}$ GeV respectively. The two measurements were based on 6 and 4 events in the channel $W \to e\nu$ respectively. Subsequently, measurements of the $W$ mass based on larger and larger datasamples were performed at LEP, Tevatron and LHC with improved techniques with respect to the past. Nowadays, thanks to the high luminosity delivered by the LHC, the data-sets are large enough for reaching the precision of $1 \times 10^{-4}$, that is the theoretical accuracy in the prediction from the EW fit, as shown in chapter 1 and in Ref. [1]: this requires to achieve an unprecedented understanding of the detector, the analysis, and the theoretical predictions.

The most precise measurements of the $W$ mass performed so far are schematically shown in figure 3.1. The most precise single measurements are the ATLAS and the CDF ones, which reached a precision of 19 MeV. The precision of the single measurements is not yet comparable with the precision reached by the EW fit of 8 MeV.
Chapter 3. Measuring the $W$ mass

Figure 3.1: $W$ mass measurements performed by LEP (all the four experiments), Tevatron ($CDF + DØ$) and LHC (ATLAS) compared to the precision of the prediction of the EW fit (Ref. [1]). The precision reached by the ATLAS and CDF measurements is 19 MeV in both cases, and these represents the most precise measurements of the $W$ mass performed up to now. This plot is taken from Ref. [26].

3.1.1 W mass measurements at LEP II

The Large Electron Positron collider was built at CERN in the 80’s for colliding electrons and positrons at center-of-mass energies ranging from the $Z$ mass (91.2 GeV) to more than 200 GeV. Its main goal was to study with high precision the EW physics, and, in fact, for precision measurements like the mass of the $Z$ mass, its precision is still unattainable with the new machines. The second LEP run, from 1996 to 2000, scanned center-of-mass energies between 130 and 209 GeV, in order to study, among others, $W$ boson and its production process. In figure 3.2, it is possible to see the two Feynman diagrams that contribute at tree level to the WW production: the measured cross section agrees with the predicted one taking into account both diagrams.

At LEP two methods for measuring the $W$ mass have been used:

- Fitting the shape of the WW production cross section as a function of the center-of-mass

\footnote{LEP II was importantly exploited for studying various physical processes: specifically, searching for the Higgs boson and for supersymmetric particles were two important goals.}
energy around the threshold $2M_W$. The left plot of figure 3.3 shows the cross section for three different values of the $W$ mass.

- Fitting the shape of the invariant mass distribution for $WW \rightarrow q\bar{q}q\bar{q}$ or $WW \rightarrow q\bar{q}l\bar{\nu}$. The distribution of the invariant mass of the reconstructed $W$ bosons in the four-jets channel, recorded by the ALEPH experiment, is shown in the right plot of figure 3.3. In the second one the neutrino can be precisely reconstructed by imposing the conservation of the initial four momentum.

An importance concept for the measurement that it is introduced here is the scale precision. In fact, the main ingredient for a precise $M_W$ measurement is the precision with which it is possible to assess that a GeV as measured by the detector corresponds to a true GeV. This is important for what concerns the main object with which the measurement is performed. As known, it is difficult to reach good scale precision with the hadronic jets, whereas it is easier for the leptons. At LEP, for what concerns the measurement with the invariant mass in the 4-jets channels, the constraint on the 4-momentum due to the knowledge of the initial state, which is very well known, sets the scale of the energy. Thus the analysis depends only on the measurement of the jet directions, removing the problem of the scale precision. In the second case with one of the $W$ decays leptonically, the neutrino can be reconstructed imposing again the 4-momentum conservation. The lepton scale is very well calibrated, producing a precise measurement in this channel too.

Only a small fraction of the luminosity has been collected near the threshold, because data at higher energy could be used to study in more detail the $W$ boson, since the cross section at higher energy is larger, and to search for the Higgs boson or for new physics.

The final LEP result, after combining the measurements performed by the four experiments, for the extraction of the $W$ mass from the cross section is:

$$m_W^{thr} = 80.42 \pm 0.20 \pm 0.03\text{(energy)} \text{ GeV}$$

where the second uncertainty refers to the energy calibration.

The measurement with the fit to the invariant mass distribution was by far more precise because of the much smaller statistical uncertainty, and the final result, again after combining the measurements performed by all the four experiments, is

$$m_W^{direct} = 80.375 \pm 0.025\text{(stat.)} \pm 0.022\text{(syst.) GeV}$$

The total uncertainty is 34 MeV, reduced to 33 MeV if the measurement from the threshold scan is included. The main systematic uncertainties are due to: LEP centre-of-mass energy, detector effects, fragmentation and hadronization, QED/Electroweak uncertainties, final state interactions, and colour reconnection.

A comprehensive report of the results of LEP II, focusing particularly on the $W$ boson properties and on the combination of the results of the various experiments, is presented in Ref. [27].
Chapter 3. Measuring the $W$ mass

Figure 3.3: Left plot: cross section for the production of a $WW$ pair at LEP as a function of the center-of-mass energy. The characteristic turn-on around twice the $W$ mass is illustrated for three different values of the $W$ mass. Figure taken from Ref. [28].

Right plot: invariant mass of the reconstructed $W$ bosons from events tagged with a four-jets final state. The data shown in this histogram was collected by the ALEPH experiment. Figure taken from Ref. [29]

3.1.2 $W$ mass measurements at Tevatron

![Feynman diagram]

Figure 3.4: $W$ production at hadron colliders, followed by a leptonic decay. The Feynman diagram represents only the hard process that involves quarks, the rest of the proton can be factorized from this hard process.

Tevatron was a proton-antiproton collider located at Fermilab, designed for collisions at center-of-mass energy of 1.96 TeV; it was operating between 1983 and 2011 with pauses for upgrades, and was the main instrument for important discoveries like the top quark and the $B_s$ oscillations.

Tevatron was the first hadron collider capable of producing a very large amount of $W$ bosons, allowing a precise measurement of the $W$ mass. Specifically, the measurement was performed looking at events with $p \bar{p} \rightarrow W + X$ with leptonic decays of the $W$ into an electron or a muon, and then fitting some kinematic distributions sensitive to the $W$ mass, such as the lepton transverse momentum, the missing transverse energy, and the transverse mass (more details about the strategy for the measurement of the $W$ mass at hadron colliders are given in the following section). The Feynman diagram related to the process under study is sketched in figure 3.4.

In this case, the leptonic decay channel has been studied, because only in this channel a very good precision scale can be achieved. In fact, in the hadronic collisions it is not possible to impose a constraint on the energy of the initial state, and calibration must be performed in a
different way.
The final measurements of the $W$ mass performed by the two experiments show that the amount of collected $W$ events were large enough to have statistical uncertainties at the level of the systematic uncertainties.

The latest CDF measurement has been performed with $2.2 \text{ fb}^{-1}$ of integrated luminosity in both electron and muon decay channels with a total of $1.1 \times 10^6$ $W$ candidates. The value of the $W$ boson mass measured by CDF is $M_W = 80.387 \pm 0.019 \text{ GeV}$. The final CDF measurement is still the most precise measurement ever performed together with the ATLAS one, that will be described in the following section.

The DØ measurement is performed with $4.3 \text{ fb}^{-1}$ of integrated luminosity in the electron decay channel with a data set of $1.68 \times 10^6$ $W$ candidates. The value of the $W$ boson mass measured by DØ is $M_W = 80.375 \pm 0.023 \text{ GeV}$ when combined with the previously analyzed $1 \text{ fb}^{-1}$ of integrated luminosity.

The Tevatron combination of these two measurement gives:

$$M_W = 80.385 \pm 0.016 \text{ GeV}$$

A complete review of the Tevatron results is given in Ref. [30], where both the experimental techniques and the implications on the SM predictions are discussed.

3.1.3 The first attempts at the LHC: ATLAS and CMS

During the first years of operations of the LHC, the ATLAS and CMS collaborations put a lot of efforts in the search for the Higgs boson, leading to its discovery in 2012, and for direct production of new physics particles. However, given the better and better knowledge of the detector and the large data-sets already collected, the two collaborations started an effort of measuring the $W$ mass.

The measurements are performed again as they were at Tevatron, with more details in the following sections. New challenges at the LHC are due to the different environment of having a proton-proton machine and a large number of pileup collisions produced in each bunch crossing. Particularly, the CMS collaboration made public a preliminary result [31] in 2016, in which the mass of the $Z$ boson was measured in a way that resembles a measurement of the $W$ boson mass. More details are given in the following sections. This is a control check that can be done for validating the analysis before measuring the $W$ mass. This exercise showed that the main limitation of the analysis was the small size of the sample on which the calibrations were derived. In this case the statistical effect of the sample for the calibrations hid possible more complex systematic effects. However, to perform the analysis on the $W$, there is something else that must be taken into account: calibrating the physical objects between $Z$ data and MC is straightforward, but applying these corrections on the $W$ requires more care. In addition the background, that is almost negligible for $Z$ events, is larger in $W$ boson events, and must be
Chapter 3. Measuring the $W$ mass

studied with attention.

The ATLAS collaboration has presented in the same year its first $W$ mass measurement [26]. The measurement is performed on the 7 TeV data-set, corresponding to $4.6 \times fb^{-1}$ of integrated luminosity. The selected data sample consists of 7.8 million candidates in the muon channel and 5.9 million candidates in the electron channel. The final result obtained by the ATLAS collaboration is:

$$M_W = 80.370 \pm 0.019 \text{ GeV}$$

that is in agreement with the previous measurements within the uncertainty. It reached the same level of precision of the CDF result. The new world average, after the ATLAS measurement, has not yet been computed. This is due to the fact that this measurement has not been published yet, and because the correlation between the systematic uncertainties of the various measurement is non-trivial to estimate.

Some of the features of these measurements will be discussed in the next sections in the context of the systematic uncertainties.

3.2 The $W$ mass measurement at hadron colliders

Usually, the mass of a particle is extracted from the reconstructed distribution of the invariant mass of its decay products. For the $W$ boson, this is extremely difficult to do at a hadron collider because:

- The visible final state ($W \rightarrow q\bar{q}'$, BR $\simeq 70\%$) appears in the detector as a pair of jets. Unfortunately this final state is produced with much higher rate in QCD inelastic processes. In addition the resolution on the dijet invariant mass is poor in comparison to the natural width of the $W$ boson. As a result, the $W$ peak in the dijet invariant mass spectrum is covered by the QCD background, making this measurement difficult to achieve. Nowadays, the peak of the dijet invariant mass produced by the decay of the vector bosons has been observed in a boosted topology. However, the resolution is so poor that it is impossible to distinguish between the $Z$ and the $W$. It is important to notice that, even in the case where a pure sample of hadronic decays of the $W$ boson was collected, a precise measurement at the level of $10^{-4}$ would be extremely difficult, because of the scale precision. In fact hadronic jets are difficult to calibrate, and the reachable precision is nowadays at the percent level. In figure 3.5, it is possible to see the di-jet invariant mass spectrum, for transverse momentum of the jet pair between 700 and 800 GeV. In this case the peak of the vector bosons is present, but $W$ ans $Z$ contributions cannot be resolved.

- The leptonic final state ($W \rightarrow l\nu_l$, with $l = e, \mu$, BR $\simeq 20\%$) has instead a very clear signature in the detector, with one high energy isolated lepton (that can be measured with high precision), accompanied by missing transverse energy. In the $\tau$ case, the lepton cannot be reconstructed with the required precision. Therefore the final states with an electron
or a muon are the most suitable for this measurement. Despite the significant signal-over-background ratio, it is not possible to reconstruct the invariant mass. Indeed the neutrino transverse momentum can only be measured by subtraction and the longitudinal component of its momentum is not accessible. In addition, because of the relatively large cross section for the $W$ production at the hadron colliders, the number of the expected $W$ bosons with lepton decays is very large, making the systematic uncertainties dominant in the analysis.

In the following, only measurements of the $W$ mass with the final state with a lepton (that can refer to an electron or a muon) will be considered. Since the invariant mass cannot be reconstructed directly, the analysis is performed by measuring experimental observables that are sensitive to the mass of the $W$ boson, whose distributions is then fitted with a template fit procedure. The template fit consists in simulating the expected distributions of these observables for different $W$ mass hypotheses, and using them as probability density functions for a likelihood scan. Since systematic uncertainties arise from discrepancy between the templates and the data, this procedure relies heavily on the MC simulation. Systematic uncertainties can be associated to mismodeling of the $W$ production and decay, of the hadronization of coulerd particles produced in the event, as well as discrepancies in the simulation of the detector effects and reconstruction.
3.2.1 Lepton transverse momentum and transverse mass

The process under study is the following:

\[ p + p \rightarrow W + X \quad \text{and} \quad W \rightarrow l\nu \]

where X, called the "hadronic recoil", is the set of all other particles produced in the collision, typically from the hadronization of the proton remnants and of the gluon initial state radiation. At the generator level, meaning with no reconstruction or pileup effects, the recoil momentum is equal and opposite to the \( W \) one. The \( W \) boson then decays into a lepton and a neutrino: the former is well measured, whereas the latter cannot be detected. The two main experimental variables sensitive to the \( W \) mass are the lepton transverse momentum and the transverse mass.

To explain the reason why these variables are useful, we can proceed by steps. To start with, suppose that the \( W \) boson is produced with no transverse momentum: in this case, in its rest frame, the lepton energy is half of the \( W \) mass, \( \frac{W}{2} \) and the lepton transverse momentum is related with its decay angle; when the decay is orthogonal with respect to the beam-line the lepton transverse momentum reaches its maximum value, equal to half of the \( W \) mass. But since the lepton transverse momentum is invariant under boosts along the beam axis, a measurement of the distribution of the lepton transverse momentum gives a measurement of the \( W \) mass.

![Distribution of muon transverse momentum in W events](image1)

![Distribution of transverse mass in W events](image2)

Figure 3.6: Left plot: Lepton transverse momentum distribution: the two cases of absent (present) \( W-p_T \) spectrum are plotted in black (blue). The effect of the \( W-p_T \) is basically to smear the Jacobian peak at half of the \( W \) mass. The green line, superimposed on the peak, indicates \( \frac{W}{2} \).

Right plot: Transverse mass square distribution. In black (blue) the distribution is shown in case of absent (present) \( W-p_T \). The two distributions look pretty similar, confirming the weak dependence of the transverse mass on the \( W-p_T \) spectrum. The green line, superimposed on the peak, indicates \( \frac{W}{2} \). The red histogram shows the effect of choosing only the charged tracks inside of the detector acceptance for computing the hadronic recoil.

Both plots show events simulated with Pythia8 standalone (for further details about this software see Ref. [33]), with no selection applied on the events

\[ 2 \text{This is exactly true only in the assumption that the lepton mass can be neglected, that is perfectly verified.} \]

The correction to the transverse momentum due to the lepton mass scales like \( \frac{m_l^2}{m_W^2} \). In the case of the muon this correction is larger than the electron, and is of the order of \( 10^{-6} \), much smaller than the target precision for this measurement.
3.2 The $W$ Mass Measurement at Hadron Colliders

The left plot of figure 3.6 shows the distribution of the lepton transverse momentum in simulated events, with no selection applied. Although the actual distributions might be slightly different, the two plots are very illustrative for highlighting the features of the variables used. The black histogram is the distribution in case of no $W$-$p_T$, and the features previously described are visible. Things are more complicated because the $W$ transverse momentum is different from zero. Its distribution, shown in figure 3.7, has a most probable value of the around 4 GeV, and an average value of the order of 20 GeV.

As a result, the distribution of the lepton transverse momentum is smeared, as it can be seen in the blue histogram in the left plot of figure 3.6. However, a measurement of the $W$ mass is still possible, provided an exact knowledge of the $W$ transverse momentum distribution. This distribution is not theoretically known with the required precision. It is possible to mitigate this uncertainty by measuring the transverse momentum of the recoil on an event-by-event basis, equal by definition to the $W$ transverse momentum, and computing the transverse mass, that is the invariant mass in the transverse plane. Starting from the invariant mass $M$:

$$M^2 = 2p lp\nu (\cosh(\Delta \eta) - \cos(\Delta \phi))$$

where $p_l$ and $p\nu$ are the transverse momenta of the lepton and the neutrino, respectively, and $\Delta \eta$ and $\Delta \phi$ represent the difference in pseudorapidity and azimuthal angle between the lepton and the neutrino, and putting $\Delta \eta = 0$, treating all the decays as they occurred completely in the transverse plane, we obtain the transverse mass $M_T$:

$$M_T^2 = 2p lp\nu (1 - \cos(\Delta \phi))$$

Since the transverse momentum is measured as missing transverse momentum, as the opposite
of the vectorial sum of the lepton and the rest, it is useful to rewrite $M_T$ in the following way:

$$M_T^2 = 2(p_l |\vec{p}_l + \vec{h}| + p_T^2 + \vec{p}_h \vec{h})$$

where we have defined the hadronic recoil $\vec{h}$ as the 2D vectorial sum in the transverse plane of the momenta of all the other particles produced in the collision, and $\vec{p}_l$ is the 2D vector related to the lepton transverse momentum. By definition, the hadronic recoil is opposite to the $W - p_T$, therefore measuring the recoil means measuring the $W - p_T$ on an event-by-event basis. Being the main topic studied in this thesis, a detailed discussion on the experimental definition of the recoil is postponed to the next chapters.

The transverse mass has, similarly to the lepton transverse momentum, its maximum at the value of the $W$ mass and, in the limit in which the $W$ transverse momentum is zero, it is equal to twice the lepton transverse momentum. Whereas the lepton transverse momentum depends linearly on $\beta_T$, that is the $W$ velocity in the transverse plane, the transverse mass depends on $\beta_T^2$ and therefore it is less affected by the $W - p_T$ spectrum. The transverse mass coincides with the invariant mass only when the decay happens in the transverse plane in the $W$ rest frame: in the other cases the transverse mass is underestimating the invariant mass. This implies that the most sensitive part of the transverse mass spectrum is in the peak, whereas the left tail heavily depends on the production mechanism, particularly on the polarization of the $W$ and on the $W - p_T$ spectrum. A plot of the transverse mass is shown in the right plot of figure 3.6, in simulated events, with no selection applied. The black and blue histograms, corresponding to the cases in which the $W - p_T$ is absent or included, respectively, are very similar, confirming the previous statement. The red histogram is produced using only the charged tracks inside the detector acceptance to compute the recoil. The poor resolution of this recoil definition degrades the peak, making it less powerful. This is only an experimental effect though, and it can be faced by choosing a better experimental definition of the recoil. Being the main topic of this thesis, it will be better discussed in the next chapters.

The potentiality of the transverse mass for a precise measurement of the $W$ mass has been already highlighted in [34] just after the $W$ boson discovery. Indeed the transverse mass can be seen as a singularity variable, obtained by considering a general variable of the lepton and recoil transverse momentum and their relative angle, and minimizing its dependence on the $W$ transverse momentum, as shown in [35]. In the course of this thesis, the variable $M_T^2$ will be always used instead of $M_T$: this is because it can be decomposed in a sum of terms, each of them can be classified for each dependence on $p_l$ and $h$. Taking the square root instead mixes the various contributions.

### 3.2.2 Systematic uncertainties

Since the LHC is delivering a great amount of integrated luminosity that reflects in several hundreds of millions of $W$ events, the systematic uncertainties become more and more relevant in this analysis. A general overview of the main systematic uncertainties is given in the following.
3.2 The W mass measurement at hadron colliders

However, the treatment of the systematic uncertainties related to the recoil is one of the main part of this work: therefore, a more accurate discussion on this side is given in the next chapters of this thesis.

In order to better analyze the systematic uncertainties, it is useful to follow the break down of the uncertainties proposed in Ref. [36].

Table 3.1: Break down of the uncertainties in the W mass measurement in CDF and ATLAS analyses. All the numbers are expressed in MeV. Table taken from Ref. [36].

<table>
<thead>
<tr>
<th></th>
<th>CDF/$p_T$</th>
<th>CDF/$M_T$</th>
<th>ATLAS/$p_T$</th>
<th>ATLAS/$M_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>16</td>
<td>15</td>
<td>7.2</td>
<td>9.6</td>
</tr>
<tr>
<td>Lepton Scale and Resolution</td>
<td>7</td>
<td>7</td>
<td>6.5</td>
<td>6.5</td>
</tr>
<tr>
<td>Recoil Scale and Resolution</td>
<td>5.5</td>
<td>6</td>
<td>2.5</td>
<td>13</td>
</tr>
<tr>
<td>Backgrounds</td>
<td>4</td>
<td>3.5</td>
<td>4.6</td>
<td>8.3</td>
</tr>
<tr>
<td>PDFs</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>10.2</td>
</tr>
<tr>
<td>W transverse momentum model / QCD</td>
<td>9</td>
<td>3</td>
<td>8.3</td>
<td>9.6</td>
</tr>
<tr>
<td>Photon Radiation/EWK</td>
<td>4</td>
<td>4</td>
<td>5.7</td>
<td>3.4</td>
</tr>
</tbody>
</table>

The systematic uncertainties can be split into two terms. First, the experimental uncertainties are those related to a non perfect calibration of the physical objects, namely the lepton and the recoil, and to the background estimation. Second, the physical modeling, that is the prediction of the $W$-$p_T$, $Y$ and polarization distributions, as well as the EW contributions to the propagators and to the final state radiation. For both the CDF and the ATLAS measurements the experimental uncertainties are subleading with respect to the theory uncertainty, although the calibration of the recoil for the ATLAS transverse mass fit was important. For what concerns the physical modeling, the biggest uncertainties arise from the knowledge of the parton distribution functions, that determines both the rapidity and the polarization, and from the $p_T$ spectrum, produced primarily by QCD initial state radiation. The EW corrections play a smaller role with respect to the rest.

Lepton transverse momentum scale

Recalling the concept of the scale precision, the lepton transverse momentum scale is of extreme importance: it is in fact the main object with which the analysis is performed. Any discrepancy between the data and the simulation in the scale reflects directly on the $W$ mass measurement: this puts a strong motivation in calibrating this scale with a precision better than the target precision on the $W$ mass. For the muons, this goal is achieved by considering muons with defined invariant mass, coming from narrow resonances such as the $J/\psi$ and the $\Upsilon$, and fitting the correction to the track curvature in the transverse plane, that is proportional to $p_T^{-1}$, due to various detector effects, such as the in-homogeneity of the magnetic field, the energy loss by passing through the tracker material and the residual misalignment of the tracker. The fit is performed in order to reproduce in the simulation the peak position of the resonances that are observed in the data. As a closure plot for validating the calibration, the $Z$ peak reconstructed with the two muons is compared between data and MC: possible residual discrepancy allows to
assess this systematic uncertainty. For the electrons the scale is calibrated through the comparison of the Z invariant mass peak between data and MC. And specific corrections are derived in order to adjust the response and the resolution between data and simulation. The resulting systematic uncertainty is usually larger for the electrons than for the muons, as the simulation of the electrons, which concerns Bremsstrahlung radiation and calorimeter simulation is more complex than the one for muons, which is mainly related to the tracker.

Recoil fragmentation discrepancy

Another important source of systematic uncertainty is the possible discrepancy between data and MC in the fragmentation as well as in the reconstruction of the recoil. The effects that contribute are, at a fixed value of the momentum of the W boson, the simulation of the underlying event and of the QCD radiation, the simulation of detector effects for the recoil (acceptance, efficiency, etc.) and the simulation of the pile-up collisions. An important step that should be done to mitigate the effect of this discrepancy is to perform the measurement in a region of small $p_T$ of the recoil, where the contribution of the lepton to the measurement is dominant, so that the effect of a possible discrepancy is limited below the target precision. Furthermore, it is important to correct the distribution of the recoil that enters in the template, in order to be as similar as possible to that one observed in the data. Since this cannot be done directly in W data, it can be done only by studying the difference between the Z MC and data samples, as better explained in the following subsection. This discussion is postponed to the next chapters, as it is one of the main topics that are dealt with in this thesis.

Physical modeling of the W kinematics

With physical modeling of the W process we refer to the prediction of the kinematic distribution of the W boson. The physical modeling is one of the biggest contributions to the total systematic uncertainty.

The differential cross section in $p_T$, $Y$ of the W boson, and $\theta$ and $\phi$ of the charged lepton in the Collins-Soper frame 3 (Ref. [37]) can be written as:

$$\frac{d\sigma}{dp_T^2 dY d\cos \theta d\phi} = \frac{3}{16\pi} \frac{d\sigma^{U+L}}{dp_T^2 dY} \left( 1 + \cos^2 \theta + \sum_{i=0}^{7} A_i(p_T,Y) P_i(\cos \theta, \phi) \right)$$

where the $d\sigma^{U+L}$ is the unpolarized cross section, and the rest describes the angular distribution of the lepton. The term $1+\cos^2 \theta$ describes the unpolarized case with no longitudinal component of the W polarization, whereas the eight coefficients $A_i$ are the so called angular coefficients. These coefficients describe the polarization of the W for each value of $p_T$ and $Y$, that eventually

---

3The Collins-Soper frame is a frame in which the W is at rest, and the z-axis is defined as that one that bisects the angle between one proton and minus the second proton directions. The x axis is defined in such way that the x-z plane contains the beam axis.
determines the distribution of the lepton. At small values of $W-p_T$, all of them are almost negligible but $A_4$, which describes the transverse polarization of the boson. More details on the angular coefficients can be found in Ref. [38].

As a first approximation one can decompose the differential $p_T$ and $Y$ distribution as the product of the $Y$ distribution times the $p_T$ distribution for a given $Y$ value. This decomposition is motivated by the fact that the $Y$ and $p_T$ distributions are almost uncorrelated, except in the region close to the edge of the rapidity distribution (where one of the initial partons has $x \approx 1$), where the $p_T$ distribution goes to lower and lower values. This statement can be proven by looking at figure 3.8, where a profile plot of the $W-p_T$ as a function of the rapidity is shown. Since the finite acceptance of the detector, the rapidity distribution proven with the analysis is far from the edge, and thus the approximation of uncorrelated variables is fairly good. This is not exactly true, but is useful to try to decorrelate the uncertainty due to a mismodeling of the $W-p_T$ spectrum form the uncertainty due to the knowledge of the pdf (that reflects in an uncertainty on the rapidity distribution).

![Figure 3.8: Profile plot of the $W^+$ and $W^-$ transverse momentum as a function of its rapidity $Y$, at generator level, with no cuts due to detector effects or event selection. The events are generated with Pythia8 standalone (Ref. [33]). The mean is shown, together with the statistical uncertainty due to the finite MC sample that is negligible. The behaviour is similar for the two charges, and it shows a flat distribution for central values of the rapidity, that goes to zero while approaching the edge of the rapidity distribution: in this case, one of the two quarks has $x \approx 1$, and there is no energy available to produce $W$ transverse momentum.](image)

**W rapidity and parton distribution functions** The EW process have been very well studied in the past and the predictions are extremely good in describing the $W$ leptonic decay, once the $W$ kinematics and its polarization are assigned. In the approximation where $W-p_T$ can be neglected with respect to $p_z$, verified as long as the measurement is performed in a region with small $W-p_T$, we can consider the beam axis as the direction of motion of the $W$ and look at the polarization along this axis. In the ultra-relativistic limit, perfectly valid for the quarks inside the protons, the scattering amplitude for a longitudinally polarized $W$ is zero. So we can neglect
Chapter 3. Measuring the $W$ mass

this state and consider only the two transverse polarizations, that corresponds to the states $m = 1, -1$ along the beam axis. Eventually, the $W$ polarization is fixed by the production flavours and directions of the incoming quarks, because of the parity violation of the $W$-fermions interaction. There is thus a correlation between the polarization of the $W$ and the rapidity distribution, that passes through the parton distribution functions, and its prediction relies on the knowledge of the pdf for the various flavours. Specifically, the polarization as a function of rapidity modifies the distribution of the lepton pseudorapidity as a function of the lepton transverse momentum. Hence the uncertainty arises from the finite acceptance of the detector, where, depending on the rapidity and polarization distributions, the number of leptons that are lost outside of the acceptance varies. With the latest measurement, the ATLAS collaboration has shown that a simultaneous fit in categories of $\eta$ of the lepton helps in reducing the uncertainty related to the knowledge of the parton distribution functions. Furthermore, looking at kinematic distributions such as the two dimensional distribution of transverse momentum and pseudorapidity of the lepton can help in further reducing this systematic uncertainty. As an example, figure 3.9 shows this two dimensional distribution for $W^-$ with with negative (left) and positive (right) helicity. The distributions significantly differ, particularly for the absence of events close to $\eta \simeq 0$ in the left plot and on the presence of these events in the right one. This allows the disentanglement of the contributions from different rapidity and helicity. This topic has been analyzed in detail in Ref. [39].

Figure 3.9: $p_T$ vs $\eta$ distribution for leptons from $W^-$ decay with negative helicity in the left plot, and positive helicity in the right plot. The differences in this distribution between the two helicity states may open the possibility to a precision measurement of the rapidity distribution for the two helicity states, as better discussed in Ref. [39]. The plots are taken from this reference.

**$W$ transverse momentum spectrum** The $W$ transverse momentum spectrum is an important ingredient of the analysis. As shown in figure 3.7, although it peaks at around 4 GeV, it has a long tail that sets the mean of the distribution around 20 GeV. As previously explained,
it modifies the lepton transverse momentum distribution in a significant way. The origin of this spectrum is the QCD initial state radiation, that makes the $W$ recoiling against the products of the hadronization. Since the bulk of the events is at small $W$-$p_T$, theoretical predictions show big uncertainties due to QCD long distance (non perturbative) effects. Moreover, in the first ATLAS measurement, it has been shown that theoretical calculations with re-summation (Ref. [40]) at NLO cannot reproduce the observed $Z$ transverse momentum distribution in data, and therefore cannot be used for predicting the $W$ transverse momentum distribution. This is in contrast to the CDF measurement, which showed an agreement between the theoretical prediction with re-summation and the data. The only differences between the two cases are the initial state ($p\bar{p}$ for CDF and $pp$ for ATLAS) and the center of mass energy (2 TeV for CDF and 7 TeV for ATLAS), which both do not appear to be relevant in terms of explaining this difference from first principles. This is still an open question that should be answered by the scientific community.

The strategy used by ATLAS was therefore to obtain the $W$-$p_T$ distribution from Pythia8, a MC simulation of the tree level process only, in which the $p_T$ is given to the $W$ boson via parton showering, more details can be found in Ref. [33], and tune the shower parameters in order to predict the $Z$ transverse momentum spectrum correctly. The $W$ transverse momentum spectrum predicted by Pythia8 after this procedure is then used for the template fit procedure. Since the difference between the LO and NLO prediction are relatively large, of the order of 5-10% in the low $W$-$p_T$ region, a discussions are still ongoing among experimentalists and theorists in order to understand the origin of this problem. Hence, the systematic uncertainties due to the imperfect knowledge of the $W$-$p_T$ spectrum is expected to be both important and difficult to properly estimate. From an experimental point of view, this systematic uncertainty can be mitigated only by a measurement of the hadronic recoil (equal to the $W$-$p_T$) on an event-by-event basis, since the transverse mass with a better and better definition of the recoil is less and less dependent on the $W$-$p_T$ spectrum. As this is one of the main topics that are studied in the course of this thesis, the discussion will be deepened in the next chapters.

QED corrections

The dominant source of electroweak corrections to $W$ boson production originates from QED final-state radiation, that must carefully simulated. Other effects, that contribute to the systematic uncertainty due to QED corrections come from the interference between ISR and FSR QED corrections (IFI), electroweak corrections due to virtual-loop and box diagrams, and final-state emission of lepton pairs.

Background estimation

In this analysis the background is typically of the order of few percent. To illustrate better the various contributions, the background percentages in the selected sample for the the analysis of the $W$ differential cross section and charge asymmetry done by CMS with the 8 TeV dataset, Ref. [41], are shown. This analysis requested one and only one isolated muon with $p_T > 24$ GeV.
(trigger threshold).

The main background contributions in this analysis are represented by (in order of importance):

- QCD multijet production, where leptons can be produced from semileptonic decays of bottom and charm hadrons, that represents about 10% of the selected sample.

- Dimuon Drell Yan production, where one lepton is missed, for instance out of the acceptance, about 5% of the selected events.

- $W$ boson decaying in a $\tau$ lepton, that can subsequently decays into a muon or an electron. This background is about 2.6% of the selected sample.

- $\tau$ pair, top quark pair, single top, and diboson production with leptonic decay are smaller, of the order of 1% of the selected events.

The various contributions are shown in figure 3.10, where the missing transverse momentum, computed as the sum of all the particles reconstructed in the event, is shown for central muons.

![Figure 3.10: Missing transverse momentum distribution in events selected with one and only one isolated muon with $p_T > 24$ GeV and $|\eta| < 0.2$. The missing transverse momentum $E_T$ is computed summing up all the particles reconstructed in the event. The $W$ production is the main contribution, but the background is important (around 20% of the events). A variable like $E_T$ is able to separate the contributions from the background: typically in $W$ mass analyses a lower bound on $E_T$ is required, in order to obtain a purer sample. This plot taken from [41].]

The background can be removed by applying proper kinematic cuts, for instance on the missing transverse momentum. The residual backgrounds can be estimated either with the simulation, particularly for EW processes, or from dedicated control region in data, specifically for QCD
background. Systematic uncertainties arise both from the normalization and from the shape of the predicted backgrounds.

3.2.3 Z events to calibrate and validate the analysis

Since the analysis relies heavily on the MC simulation, it is important to calibrate and check the consistency of the simulation and to assess systematic uncertainties using real data. An important tool for this purpose are $Z$ events decaying into two leptons. These events are produced with great abundance at the LHC: taking into account the different branching fractions into electrons and muons, and the smaller production cross section, the $Z$ events are about one tenth of the $W$ ones. They can also be identified with extremely high purity, about 99.9%, by requiring a tight lepton identification together with the $Z$ mass constraint. These events are very similar to the $W$ ones, where the main difference is that two leptons are present in the final state instead of a lepton and a neutrino: therefore they can be used to correct discrepancy between data and MC simulation and to assess some of the systematic uncertainties. In addition the production mechanism, the kinematics, the polarization and of course the mass are different between the $W$ and the $Z$ boson cases: extrapolating and propagating calibration or correction between $Z$ and $W$ requires therefore attention and specific systematic uncertainties should be quoted.

For instance this tool can be used to check whether the lepton transverse momentum scale in MC agrees with data in a momentum regime that is fairly close to the one of the $W$ mass analysis. Additionally, concerning the hadronic recoil, any discrepancy between the MC and the data due to hadronization or detector effects can be corrected on $Z$ events. These corrections can then be propagated to the case of $W$ production.

Eventually, by removing one of the two leptons, the event can be treated as $W$-like event, because the final state looks like the $W$ one; therefore the analysis designed for the $W$ events can be applied to the $W$-like events as well. Hence, a final crosscheck for validating the analysis is to measure the $Z$ mass using $W$-like events and compare it to the world average of the $W$ mass measurements, that is $m_{Z} = 91.1876 \pm 0.0021$ GeV, whose uncertainty is dominated by the LEP measurements: this should asses the consistency of all the uncertainties already quoted, and check whether new systematic uncertainties have not been taken into account.

This method will be exploited heavily throughout this thesis, therefore an in depth discussion will be given later on.

3.3 Conclusions

In this chapter, a general overview of the $W$ mass measurements has been shown, presenting the results obtained at the LEP, the Tevatron and the LHC. A detailed description of the $W$ mass measurement at hadron colliders is given, illustrating both the techniques used in the analysis and the main experimental and theoretical systematic uncertainties. Since the LHC has already collected a huge amount of luminosity, these uncertainties represent the biggest
Chapter 3. Measuring the $W$ mass

challenges towards more precise measurements of the $W$ mass.
Chapter 4

Study of $W$ production at LHC on simulated events

In this chapter several aspects of the $W$ events are analyzed in the detail, focusing particularly on the $W$ kinematics and on the features of the hadronic recoil. The effect of the events selection on the transverse mass, performed through various cuts on kinematic variables, such as the lepton $p_T$, the missing transverse energy, and the recoil $p_T$, are shown, leading to the definition of the datasets that will be used in the following analysis.

Almost all the studies performed in this chapter are made with simulated events with Pythia8 standalone [33]. $pp$ collisions with subsequent $W$ boson production have been simulated at a center-of-mass energy $\sqrt{s} = 8$ TeV, and using the NNPDF2.3 QCD+QED LO pdf [42]. Multi-parton interactions (MPI) have been simulated: they are important to properly reproduce the observed distributions for what concerns the hadronic activity. When needed, the detector effect has been simulated by imposing cuts for the acceptance in $p_T$ and $\eta$ on charged tracks, leptons and neutral particles, corresponding to the features of the CMS detector. Some of the studies and the final event selection, are instead performed with officially CMS simulated samples, containing both the simulation of the event and the detector effects. In this latter case it will be explicitly mentioned.

4.1 The $W$ production mechanism

$W$ production at hadron colliders happens predominantly via quark-antiquark interactions. The relative rate of the various quark-antiquark production channels depends on the parton distribution functions and on the parameters of the CKM matrix. This also reflects in an asymmetric production between $W^+$ and $W^-$, for what concerns both the production rates and the rapidity spectrum.

As these studies are performed with a LO MC generator, no diagrams with gluon-gluon and quark-gluon production are included. However these processes are simulated through the parton
shower in the soft and collinear regime. The importance of these diagrams depends on the factorization scale $\mu_F$, which divided the soft and collinear regime from the hard scattering one. If the computation is performed at the $n$–th order, the rate of these processes depends on $\mu_F$ at the $n+1$–th order in perturbation theory. Only in the case in which all the perturbative orders are included the result does not depend on $\mu_F$. In the conditions under study, the importance of the gluon-gluon and quark-gluon production is estimated using MADGRAPH (Ref. [43]). It is found to be of the order of 1%, negligible with respect to the larger contributions but not with the smaller ones.

The main contributions are represented by $u\bar{d}$ and $d\bar{u}$, following by the almost symmetric contributions of $c\bar{s}$ and $s\bar{c}$. The other contributions are suppressed by the out-of-diagonal elements of the CKM matrix and are thus of few percents, like $u\bar{c}$ or $d\bar{s}$, whereas the contributions of the $b$ quark is completely negligible. A summary of the various production flavours, showing both the relative fraction and some features of the corresponding $W$-{$p_T$} spectrum, whose discussion is postponed to the next section, is given in table 4.1 for $W$ bosons and in table 4.2 for $Z$ bosons.

**Table 4.1:** Summary of the relative rates and of the features of the $p_T$ spectrum for $W$ bosons, for the different quark-antiquark production channels. The mean, together with its error, is obtained by averaging on the full spectrum. The mean error is very large for the case where the number of events is very small because of the small fraction. MPV stands for the most probable value, i.e. the peak of the distribution.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$ud$</td>
<td>16.79</td>
<td>0.01</td>
<td>20.50</td>
<td>2.75</td>
<td>50.655</td>
</tr>
<tr>
<td>$d\bar{u}$</td>
<td>16.44</td>
<td>0.01</td>
<td>19.73</td>
<td>2.75</td>
<td>32.649</td>
</tr>
<tr>
<td>$c\bar{s}$</td>
<td>19.90</td>
<td>0.03</td>
<td>20.21</td>
<td>5.25</td>
<td>6.189</td>
</tr>
<tr>
<td>$s\bar{c}$</td>
<td>19.98</td>
<td>0.03</td>
<td>20.41</td>
<td>5.75</td>
<td>6.186</td>
</tr>
<tr>
<td>$u\bar{s}$</td>
<td>18.54</td>
<td>0.05</td>
<td>21.34</td>
<td>3.75</td>
<td>2.021</td>
</tr>
<tr>
<td>$d\bar{c}$</td>
<td>19.14</td>
<td>0.06</td>
<td>21.11</td>
<td>5.75</td>
<td>1.155</td>
</tr>
<tr>
<td>$s\bar{u}$</td>
<td>17.66</td>
<td>0.08</td>
<td>19.22</td>
<td>3.75</td>
<td>0.578</td>
</tr>
<tr>
<td>$c\bar{d}$</td>
<td>18.76</td>
<td>0.09</td>
<td>20.30</td>
<td>3.75</td>
<td>0.552</td>
</tr>
<tr>
<td>$b\bar{c}$</td>
<td>22.58</td>
<td>0.85</td>
<td>22.14</td>
<td>7.75</td>
<td>0.007</td>
</tr>
<tr>
<td>$c\bar{b}$</td>
<td>22.28</td>
<td>0.72</td>
<td>18.44</td>
<td>7.75</td>
<td>0.007</td>
</tr>
<tr>
<td>$u\bar{b}$</td>
<td>22.12</td>
<td>3.17</td>
<td>21.00</td>
<td>3.75</td>
<td>0.000</td>
</tr>
<tr>
<td>$b\bar{u}$</td>
<td>23.46</td>
<td>6.44</td>
<td>23.24</td>
<td>7.25</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table 4.2:** Summary of the relative rates and of the features of the $p_T$ spectrum for $Z$ bosons, for the different quark-antiquark production channels. The mean, together with its error, is obtained by averaging on the full spectrum. The mean error is very large for the case where the number of events is very small because of the small fraction. MPV stands for the most probable value, i.e. the peak of the distribution.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$u\bar{u}$</td>
<td>18.70</td>
<td>0.06</td>
<td>22.13</td>
<td>3.15</td>
<td>43.488</td>
</tr>
<tr>
<td>$d\bar{d}$</td>
<td>18.26</td>
<td>0.06</td>
<td>21.92</td>
<td>3.45</td>
<td>37.702</td>
</tr>
<tr>
<td>$s\bar{s}$</td>
<td>19.55</td>
<td>0.10</td>
<td>20.78</td>
<td>4.35</td>
<td>11.085</td>
</tr>
<tr>
<td>$c\bar{c}$</td>
<td>21.30</td>
<td>0.16</td>
<td>21.33</td>
<td>5.85</td>
<td>4.970</td>
</tr>
<tr>
<td>$b\bar{b}$</td>
<td>25.86</td>
<td>0.22</td>
<td>21.84</td>
<td>10.65</td>
<td>2.755</td>
</tr>
</tbody>
</table>
4.1 The W production mechanism

4.1.1 The origin of the $W$-$p_T$

The $W$-$p_T$ spectrum peaks around 5 GeV and presents a long tail to higher values. In a simplified framework, quarks inside the proton move colinearly to the proton, bringing no transverse momentum. Thus at tree level no transverse momentum of the boson is present. This is typically generated by initial state radiation (ISR): a quark emits a gluon with transverse momentum, making the $W$ boson recoiling against the radiation. In the region of large $p_T$, the computation can be performed using the perturbative expansion. However, this is not the case for the small $p_T$ region, where non-perturbative effects play an important role. Resummation techniques (Ref. [40]) are then needed and used in the most recent and accurate computations.

![Figure 4.1: $p_T$ spectrum for $W$ (left) and $Z$ (right) simulated events. No event selection has been applied. The plots show the normalized distributions for the different flavours participating in the production. Contributions to the total rate that are smaller than 1% are removed from the plot, in order to make the plot more readable.](image)

The two plots in figure 4.1 show the $p_T$ spectrum for $W$ (left) and $Z$ (right) simulated events, divided in categories based on the flavours participating in the production. Only generator level quantities have been studied. No kinematic cut has been applied on these events. Two interesting features are visible in these plots. The first one is that the heavier the flavours participating in the production, the harder the resulting $p_T$ spectrum. The second one is that the spectrum is almost proportional to the mass of the boson, making the $Z$ spectrum harder than the $W$ one. This is coherent with the result of the theoretical studies and QCD calculations: Ref. [44] and [45].

The left plot in figure 4.2 shows that there is an agreement between the distributions of $p_T/M$ for $W$ and $Z$ events in the low $p_T$ region. The agreement is studied for the two cases where only light flavours ($u\bar{d}$ for the $W$ and $u\bar{u}$ for the $Z$ events) or all the flavours are considered. Considering only light flavours is meant to remove the contributions of heavy quarks that is not present in the $W$ events. This agreement is quantified with the ratio plots shown in the right plot of figure 4.2. The ratio between the distributions of $p_T/M$ for $Z$ and $W$ events is shown, for the two cases of all flavours (black) and light flavours only (red). The discrepancy from the unitary ratio is of the order of 10%, showing that the predicted proportionality between $p_T$ and
Chapter 4. Study of W production at LHC on simulated events

Chapter 4. Study of W production at LHC on simulated events

Section 4.1

4.1. Study of W production at LHC on simulated events

4.1.1 Boson Pt distribution scaled for W and Z events

Figure 4.2: Left: distribution of $p_T/M$ for W and Z simulated events, for the two cases of light flavours only (green for the $W$ and blue for the $Z$ ) and all the flavours (black for the $W$ and red for the $Z$ ). The distributions agree qualitatively.

Right: ratio plots of the $W$ and $Z$ distributions. The behaviour is similar for the two cases of light flavours only (red) and all the flavours (black), and show discrepancy of the order of 10% from the unitary ratio.

$M$ is respected with a reasonable precision.

4.1.2 The rapidity spectrum

Figure 4.3: Rapidity spectrum for $W^+$ and $W^-$ with no cut applied (left) and with kinematic cuts on the lepton (right). The cut applied are $p_T > 30$ GeV and $|\eta| < 2.4$, and are applied on the generator level quantities. The cuts modify the shape of the distributions, particularly reducing the range of the rapidity spectrum.

The $W$ rapidity spectrum depends mainly on the parton distribution functions, and it is thus different for $W^+$ and $W^-$. The right plot in figure 4.3 shows the rapidity spectrum for $W^+$ (red) and $W^-$ (blue) with no selection applied. The ratio between the rates of $W^+$ and $W^-$ is about three to two. The kinematic cuts on the lepton ($p_T > 30$ GeV and $|\eta| < 2.4$, applied directly on the generator level quantity) modifies the shape of the distributions (left plot of the same figure), leaving almost unaffected the ratio between the rates of positive and negative $W$. 

48
4.2 A closer look at the recoil

The hadronic recoil is defined as the sum of all the particles produced in the collision but the boson. Its momentum is equal, by definition, to the momentum of the boson. As the transverse momentum of the $W$ is an important ingredient of the analysis, difficult to predict theoretically, the measurement of the recoil is extremely important in order to experimentally access the $W - p_T$.

It is better to consider the hadronic recoil than the MET because the MET mixes the two main objects of these events: the lepton and the recoil. For instance, the simulation of the lepton and the recoil are different and attention is paid to different aspects and problems: considering the MET would mix the systematic uncertainties of the two objects.

A sketch of a $W$ event is shown in the left plot of figure 4.4. The recoil is composed by both the remnants of the protons and the hadrons produced by the gluon ISR. At a first approximation, the $p_T$ is carried only by the hadrons produced by ISR, whereas the proton remnants do not bring any transverse momentum. However, there is no clear distinction between tracks belonging to the proton remnants and to the gluon ISR, as long range QCD effects and colour reconnection are important. Since in the bulk of the events the $p_T$ is small, these events are usually far from the jet regime, making again important to consider all the particles produced in the collision.

![Figure 4.4: Left: sketch of a $W \rightarrow l\nu$ event. The particles going to the bottom represent the hadrons produced by ISR, whereas those going close to the beam axis are meant to be the proton remnants. Right: transverse view of a simulated $W$ event with the CMS detector. The yellow line corresponds to the high $p_T$ muon coming from the $W$ decay, whereas the other particles belong to the hadronic recoil.](image)

The right plot in figure 4.4 shows a simulated $W$ event with the simulation of the detector effect. The yellow line represents the high energy muon coming form the $W$ decay, whereas the other particles are those belonging to the hadronic recoil. No pileup has been simulated. This event display clearly shows the main features of the hadronic recoil in this low $p_T$ regime: it is composed by many soft particles with no preferred direction.
The left plot in figure 4.5 shows the distribution of the number of particles for different categories, selected in $p_T$, $\eta$, and charge. The average number of particles per collision reduces a lot from around 300 in case no selection is applied to around 60 if only the charged particles in the acceptance of the tracker are considered. It is possible to conclude that most particles are produced with small $p_T$ and go forward, outside of the acceptance of the detector. Because forward particles carry typically large longitudinal momentum, the longitudinal component of the recoil is not experimentally accessible. However, as they usually carry small $p_T$, a measurement of the transverse momentum of the recoil is worth. This last statement is supported by the right plot in figure 4.5. This plot shows the profile histogram of the reconstructed recoil along the lepton direction versus the same quantity at generator level, for the same selections shown in the left plot. This quantity is of interest because the transverse mass $M_T^2$ can be expanded in series of $h/p_T^\mu$, that in this low $p_T$ regime, is a parameter of order 10%. The first order expansion is the following:

$$M_T^2 \simeq 4p_l^2 + 4\vec{p}_l \cdot \vec{h}$$

(4.1)

so the first correction containing the recoil depends only on the parallel component of the recoil itself to the lepton direction. So the component of the recoil along the lepton direction is a good estimator to judge the impact of the recoil on the $M_T$. The profile plot discussed before shows a good correlation between the gen-level and the reconstructed quantities, making interesting to study the $p_T$ of the recoil. Particularly, the blue line, representing the case where all the particles with $|\eta| < 5$ and $p_T > 300$ MeV are used to compute the reconstructed recoil, is very close to the identity functions. This shows that even if most forward particles are not measured, the correlation between the measured and the true value is very large.
4.3 Event selection

Several cuts are usually applied on $W$ events in order to reduce the background and select the most suitable phase space for the measurement. The aim is put to select events where the contribution of the lepton is more important than the recoil. In this case the first order approximation, seen in the previous section, is valid, and it tells that, qualitatively, the data/MC discrepancy of the lepton should be one order of magnitude smaller than the one related to the recoil. This condition is important as calibrating the hadronic recoil is more difficult than the lepton. Additionally it is important to select events where the decay occurs close to the transverse plane, where the transverse mass coincides with the invariant mass of the lepton-neutrino pair.

Typically only high $p_T$ leptons are chosen because of the trigger thresholds, and because the higher the momentum, the closer the decay to the transverse plane. In addition, a cut on the reconstructed recoil is put in order to mitigate the presence of events with large $W$-$p_T$. Often a cut on $E_T^{miss}$ is used in order to reject the QCD background.

4.3.1 Effect of the selection on the transverse mass

The effect of the cuts previously mentioned is illustrated with the plots in figure 4.6, which show the evolution of the transverse mass distributions in the following four situations (going from left to right): before any cut, requiring leptons in the geometrical acceptance, requiring leptons in the acceptance and with $p_T > 30$ GeV, requiring also TK recoil < 30 GeV\(^1\). Each plot shows various $M_T$ distributions, built using no recoil (black), ideal recoil (blue and red), TK recoil (purple), and the two first order approximations (green and brown). The cuts mainly affect the events on the left tail, whereas those close to the peak, that are the most sensitive to $M_W$, are less affected by the selection. This effect is visible for all the $M_T$ distributions shown in the plots.

4.3.2 Selecting the events on the data-sets

The selection of the events used in the following analyses is illustrated. From now on the events used for the analysis belong to the CMS officially simulated or collected samples. The events are taken from the 8 TeV datasets. There are 23.3 fb\(^{-1}\) of data collected by CMS at 8 TeV during 2012. The simulated samples are produced with MADGRAPH (Ref. [43]) and consist of 55.9 M $pp \rightarrow W (\ell \nu) +$Jets and of 29.7 M $pp \rightarrow$ Drell-Yan($\ell\ell$)+Jets. Only the muon channel has been studied, because it is experimentally easier and cleaner than the electron one. This does not affect in an important way the studies concerning the recoil, and it is a step necessary only for selecting the events for the following analysis. The fake rate is another important ingredient, but requiring high quality muons this effect can be neglected. All the events, both data and simulation, are required to pass the HLT_IsoMu24 path. It requests an isolated muon with $p_T$

\(^1\)The exact definition of the TK recoil will be given in the following subsection. It is basically the recoil measured as the sum of the charged tracks coming from the primary vertex. In this context it is used as a proxy for the $W$-$p_T$ : a cut on this variable reduce the events at large $W$-$p_T$. 

51
Figure 4.6: Effect of the event selection on the transverse mass distribution. Each plot shows the distributions of the transverse mass computed in various ways: generator level quantities in blue and red, TK recoil\(^2\) in purple, no recoil in black, and the two first order approximations, in green and brown. Moving from left to right, several cuts are progressively applied. From left to right: no cut, \(|\eta| < 2.4\), \(p_T > 30\) GeV, recoil \(TK < 30\) GeV. The \(M_T\) distribution is sculpted: the events on the tail are affected in a more important way than the peak, making it sharper.

larger than 24 GeV. Subsequently all the muons are required to have tightID (Ref. [46]) and their distance on the z axis from the primary vertex is required to be smaller than 1 mm. The primary vertex is selected as the vertex with the largest transverse momentum scalar sum.

Then \(W\)-like triggered muons are searched among the selected muons. A muon is called \(W\)-like triggered if:

- \(p_T^\mu > 30\) GeV and \(|\eta^\mu| < 2.1\)
- Particle Flow \(E_T^{miss}\) (PFMET) \(> 30\) GeV, where the PFMET is computed by summing up the transverse momenta of all the PF candidates
- relative isolation smaller than 0.5. The relative isolation is defined as

\[
\text{relIso} = \frac{\sum\text{tracks from PV } p_T^l + \max(0, \sum\text{neutral PF candidates } p_T^l - 0.5 \sum\text{tracks not PV } p_T^l)}{p_T^\mu}
\]

The first cut is meant to reduce the effect of the turn-on of the trigger and acceptance thresholds,
whereas the other two are mainly used to reduce background events and fake muons produced by hadrons in-fly decay.

An event is subsequently selected for the $W + (-)$ sample if one and only one muon $w$-like triggered is found and it has positive (negative) charge.

Using $Z$ events is important for two steps. The first one is to study the recoil on the data, with the knowledge of the almost true value of the recoil as reconstructed with the muon pair. The second one is to test the whole analysis on the data, by measuring the $Z$ mass a la $W$-like, fitting the transverse mass distribution. For this second purpose it is important to select $Z$ events as they were $W$ ones.

Therefore, first of all $Z$ candidate events are identified within the $Z$ simulated events or the data events passing the HLT_IsoMu24 path. $Z$ pair candidates are built from two opposite charge muons whose invariant mass is in the range 70, 110 GeV. If more than one $Z$ pair candidate is found in the same event, only the closest in mass to the nominal value is considered. The event belongs to the $Z + (-)$ sample if the positive (negative) muon is $W$-like triggered with the following caveats:

- All the thresholds in GeV are rescaled by the ratio $M_Z/M_W$
- PFMET is computed by removing the other muon, treating it as a neutrino
- The other muon is required to have $|\eta| < 2.4$, $p_T > 10$ GeV, and relIso < 0.5

The last requirement is important in order to reduce the presence of fake $Z$ events where one muon comes from the decay of a $Z$ or a $W$ boson and the second one from an in-fly decay of a charged hadron. It is important to notice that an event is possibly selected for both the $Z+$ and $Z-$ samples.

### 4.4 Impact of the experimental resolution on the transverse mass

The resolutions in measuring the two objects on the analysis, the lepton and the recoil, are two other important experimental aspects. The effect of the resolution on the $M_T$ distribution is analyzed in the following.

#### 4.4.1 The impact of the lepton on the transverse mass

The two main experimental issue in measuring the leptons are the presence of final state radiation (FSR) that cannot be trivially associated to the lepton and the experimental resolution.

Figure 4.7 shows the $M_T^2$ distribution for different muon and recoil definitions. Keeping the recoil at its generator level, it is possible to notice that the effect of the FSR is small and transforms the blue histogram (generator level muon) to the red one (muon post FSR). The effect of the resolution, that is larger than the FSR one, can be seen by looking at the case where the reconstructed muon is used (black histogram).
Figure 4.7: $M_T^2$ distribution for $W$ simulated events, using different definitions of the objects. In the legend, the definition of the objects is indicated: $m$ stands for the muon and $r$ stands for the recoil. The suffixes $MC$, $MC_{\text{postFSR}}$, RECO, TK stands for generator level, generator level post FSR, reconstructed level, tracks, respectively. $1^\text{st}$ is added to indicate that the first order approximation of the $M_T^2$ has been used.

The impact on the transverse mass of these effects is negligible with respect to the experimental resolution of the recoil, as it will be clearer in the following subsection.

### 4.4.2 The problem of the recoil definition

As it emerges from the previous section, the hadronic recoil is a complex object, whose definition is not trivial. The only information needed for the computation of the $M_T$ is composed by the transverse momentum and the azimuthal angle of the recoil. However, the recoil, as obtained by the reconstruction of the event collected by the detector consists of a set of particle flow candidates, that contains a much larger information than the two components of a transverse momentum.

The recoil definition is the function that takes as input some features of the recoil and gives as output the two components of the recoil: it concentrates into two numbers all the relevant information of the recoil for the $M_T$ computation. The features of the recoil are quantities that can be computed starting from the set of PF candidates, except the lepton. For instance, one simple definition of the recoil is the so called TK recoil, which is computed as the vectorial sum of the transverse momenta of the charged tracks coming from the primary vertex. Another minimalistic definitions is the PF recoil, which is the vectorial sum of the transverse momenta of all the particles reconstructed in the events except the lepton.

The problem of finding better and better definitions of the recoil is one of the main topics of this
4.4 Impact of the experimental resolution on the transverse mass

thesis: thus a more detailed discussion is postponed to the next chapter.

4.4.3 The impact of the recoil on the transverse mass

Whereas the lepton only slightly modifies the $M_T$ distribution, the recoil affects it largely because of the poor experimental resolution. The plot in figure 4.7 shows the $M_T^2$ distribution for $W$ events, comparing different cases. It is interesting to compare the two cases in which, keeping the muon at the reconstructed level, the true value for the recoil is used (black histogram) or the reconstructed TK recoil is plugged (brown case). The effect of the poor resolution of the recoil is to smear the peak of the distribution in an important way. The importance of this effect can be seen in comparison with the effect of the muon resolution, which brings the blue histogram, obtained with gen-level quantities, to the black one. The take-home-message is that the recoil resolution affects the $M_T$ distribution in a much more important way than the lepton resolution does.

4.4.4 Statistical sensitivity in bins of the lepton $p_T$

A study of the $M_T^2$ distribution in bins of the lepton $p_T$ (called $p_l$) has been performed (left plot of figure 4.8): fixing the value of the lepton $p_T$ makes the transverse mass dependent only on the recoil, highlighting its effect on the $M_T^2$ distribution. In the course of this section the reconstructed lepton $p_T$ is used. However it has been verified that the difference with using the generator level quantity are negligible. The recoil is instead at the generator level, in order to study the problem without mixing it with the resolution. The variable under study is actually the $M_T^2$ approximated at the first order in $h/p_l$, equation 4.1. This is done in order to better connect this study with the recoil resolution.

![Figure 4.8: Left: distribution of the first order approximation of the transverse mass in bins of the lepton $p_T$. The transverse mass is computed using the reconstructed lepton and the generator level value of the recoil. Right: zoom of the previous distribution in the right tail of the peak. The superimposed dashed lines identify the positions of the values where the distribution of $M_T^2$ in the bin with 40 GeV < $p_l$ < 41 GeV (blue histogram) takes the values corresponding to 90% and 10% of the peak value.](image-url)
As seen in the plot, the bins with lepton $p_T$ close to $M_W/2$ (the blue and the green) shows sharper $M_T^2$ distribution with respect to the other. In addition this bins peak at $M^2_W \approx 6400 \text{GeV}^2$, confirming that these events are those in which the transverse mass coincides with the invariant mass, because the decay in the $W$ rest frame occurred in the transverse plane. The events in the other peaks are instead less dependent on $M_W$: the positions of their peaks depends more on the production mechanism than on $M_W$. This puts a strong need for selecting events around the peak and with $p_l \approx M_W/2$.

The right plot of figure 4.1 shows a zoom of the previous distribution in the right tail of the peak. Considering the blue curve, which is the $p_l$ bin with the sharper peak, the width of this tail has been evaluated by considering the interval where the distribution drops from the 90% (around $6500 \text{ GeV}^2$) to the 10% (around $7100 \text{ GeV}^2$) of its peak value. We can ask which recoil resolution would smear an hypothetical delta-like peak producing this result. To answer this question we can go back to the first order approximation:

$$M_T^2 \simeq 4p_l^2 + 4p_l \cdot \vec{h} \implies \Delta M_T^2 \simeq 4p_l \Delta h$$

which implies:

$$\Delta h = \Delta M_T^2/4p_l \simeq 600 \text{ GeV}^2/(4 \cdot 40 \text{ GeV}) \simeq 4 \text{ GeV}$$

This tells that the natural width of the $M_T^2$ is equivalent to a recoil resolution of about 4 GeV: thus there it is not useful to obtain a recoil with a better resolution.

Furthermore, a study of the sensitivity to $M_W$ of the various bins in lepton $p_T$ has been performed. It has been performed with a toy version of the template fit to the $M_T^2$ distribution, which has been developed to study the systematic uncertainties, and which is explained in section 6.2. A template fit to the various bins of lepton $p_T$ has been performed, and the expected statistical uncertainty on the fit result in case of 30 millions $W+$ (it is the expected number of $W+$ present in the CMS 8 TeV dataset after the selection) has been analyzed. In the left plot of figure 4.9 the $\chi^2$ scan of the $M_W$ parameters are shown, for the various bins of $p_l$: the black curve, which is the bin with $p_l \simeq M_W/2$, is the narrower, resulting in a smaller statistical uncertainty on the $M_W$ extraction. This confirms what previously observed. The right plot of figure 4.9 shows a summary of the results in terms of statistical sensitivity versus the lepton $p_T$ value. The various curves refers to an additional Gaussian smearing of 0, 3, 6, 9, and 12 GeV on the recoil value (along the lepton direction) which emulates the effect of the detector resolution. The statistical uncertainty changes significantly among the various bins, up to a factor 2 between the lowest or higher lepton $p_T$ bins and the central one. The additional resolution modifies significantly the sensitivity: a 6 GeV resolution results in a statistical uncertainty which is about a factor two larger than the ideal case.

It would be useful to repeat this study in order to evaluate the sensitivity to the $W$-$p_T$ spectrum. In fact it is expected that the lepton $p_T$ bins closer to $M_W/2$ are less sensitive to the $W$-$p_T$ spectrum, as in these bins the transverse mass coincides with the lepton-neutrino invariant mass.
4.5 Conclusions

In this chapter the main features of the $W$ events have been illustrated, focusing on the $W-p_T$ and on the recoil description. Then, the datasets and the event selection which will be used in the following have been illustrated. We have not gone through the details of the experimental definition of the recoil, that will be treated in the detail in the next chapter.

Figure 4.9: Left: $\chi^2$ scans related to the template fits to the $M^2_T$ distributions for the different bins of the lepton $p_T$. Right: statistical uncertainty on the $M_W$ extraction, based on the $M^2_T$ template fit, as a function of the lepton $p_T$ bin. The different curves and colours refer to the $\sigma$ of the injected Gaussian recoil resolution.
Chapter 5

Towards a better recoil definition

In this chapter the problem of the recoil definition is treated in detail, taking into account all the experimental issues. The classic recoil definitions are illustrated, and a method to define an improved recoil, further exploiting the information collected by the detector is developed. Eventually, the goodness of these definitions are compared in terms of the resolution on physical quantities, and they will be better assessed in the next chapter in terms of systematic uncertainties.

5.1 Defining the recoil

The recoil is a 2D vector which lies in the transverse plane, whose components enter in the $M_T$ computation. The information collected by the detector is much bigger, and consists of a set of reconstructed PF candidates: from them a great variety of different quantities can be computed. In an abstract way, the recoil definition is the mathematical function $y = f(x)$, which takes as input several quantities computed starting from the information collected by the detector and produces as output the two components of the recoil. The input variables are called $x$ and often referred as features of the recoil, whereas the output variables, the components of the recoil, will be referred as $y$. Thus, the recoil definition concentrates in a 2D space all the information needed for the $M_T$ computation and contained in the input space, which might be a far bigger space. The inputs should be computed starting from all the reconstructed particles in the event except for the lepton. This is important in order not to correlate the two relevant objects that enter in the transverse mass computation. Eventually this leads to treat uncorrelated systematic uncertainties, as it will be better described in the next chapter. In fact the kinematic distribution of the lepton is correlated to the recoil only through the $W$ kinematics, implying that at a fixed value of the $W$ kinematics, the features of the recoil are uncorrelated with the kinematics of the lepton from the $W$ decay.

The key point is that the recoil is important to mitigate the uncertainty due to the imperfect knowledge of the $W-p_T$ spectrum, and that the recoil definition represents a measurement of the $W-p_T$ on an event-by-event basis. A better definition of the recoil makes the transverse
mass less dependent on the $W$-$p_T$ spectrum. In addition a good definition should result in a lower statistical uncertainty. Roughly, we can say that the dependence on the $W$-$p_T$ spectrum is related to the central value of the recoil, and its variance affects the statistical uncertainty.

There is no intrinsic systematic uncertainty due to the recoil definition, even if, on average, it is not measuring the $W$-$p_T$ correctly. For instance, let us suppose to use the definition that assigns the vector with magnitude zero every event to the recoil. In this case the transverse mass becomes twice the lepton $p_T$, and the recoil is wrong by 100%. However, there is no bias in extracting $M_W$ with this variable. As long as the definitions used in the data and in the simulation are the same, there is no bias related to the incorrect measurement of the recoil.

A source of systematic uncertainty related to the recoil arises from the mis-modeling of the recoil in the MC simulation: it depends on the definition, but can be corrected by properly comparing $Z$ events in data and MC and extracting appropriate corrections. This topic will be further investigated in the next chapter, in the context of the systematic uncertainties.

The problem of the experimental definition of the hadronic recoil is not trivial, since as shown before, it is composed of many forward and low $p_T$ particles. Such particles are likely to be outside of the detector acceptance, or below the reconstruction threshold. Furthermore such particles can be easily confused with particles produced in pile-up collisions.

Mismeasurement of these forward and low $p_T$ particles leads to mismeasure both the magnitude and the direction of the recoil. As a result, the recoil should be really treated as a 2D object, and the resolution as well. It is more difficult to characterize it, and, for most recoil definitions, the magnitude of the resolution is comparable with the magnitude of the recoil.

The plot in figure 5.1 is useful to better understand the importance of the measurement of the angle. It shows the $M_T^2$ distribution computed with $W+$ simulated events using the reconstructed lepton, for different bins of true boson $p_T$. The various colours are related to different recoil definitions:

- blue: generator level recoil
- green: true angle and the reconstructed magnitude
- black: true magnitude and the reconstructed angle
- red: both angle and magnitude are reconstructed

where the reconstructed recoil is computed by vectorially summing up all the momenta of the charged particles from the primary vertex, except for the lepton (this is what is called track recoil, and it will be further explained in the next section). In the region where $W$-$p_T$ is between 4 and 10 GeV, where the peak of the distribution lies, the red distribution is sharper than the blue one, meaning, qualitatively, that is better to know the angle than the magnitude. On the other hand, at larger values of $W$-$p_T$, the recoil starts to fall in the jet regime: in this case, even with the charged tracks only, the direction is well measured, whereas the reconstructed magnitude gives a bad measurement of the true magnitude of the recoil.
5.1 Defining the recoil

The two easiest and most common experimental definitions of the hadronic recoil are track recoil (TK recoil) and particle flow recoil (PF recoil).

The TK recoil is computed by vectorially summing up all the transverse momenta of the tracks coming from the primary vertex (PV), except for the lepton track. The PV is identified as the vertex with the highest $p_T$ scalar sum. The lepton is required to be compatible with this vertex. In the case of high $W-p_T$, the recoil is in the jet regime, and this definition is expected to measure only a fraction of the total momentum, with the correct direction. This is not true at small $W-p_T$, where the recoil is made of soft particles distributed almost spherically in the detector. In this case the TK recoil magnitude can even be on average larger than the true recoil, and the precision on the angle is poor. In fact, it is enough to lose just one particle out of the acceptance to modify the recoil significantly. This definition has a small dependence on the pileup since the assignment of charged tracks to the vertices is done with good accuracy, thanks to the good resolution on the impact parameter of tracks, being much smaller than the average distance between vertices. The PF recoil is instead the vectorial sum of all the transverse momenta of the reconstructed particle flow candidates in the event. Since the sum of the particle flow candidates coming from pileup vertices is zero at generator level, this computation should result in the $W$ hadronic recoil. However, this condition is only true at generator level, and does
Chapter 5. Towards a better recoil definition

not takes into account the acceptance and resolution of the detector. This leads to a dependence of the PF recoil resolution as a function of the number of pile-up vertices. The resolution on the PF recoil is expected to be proportional to the square root of the number of pileup vertices, making this definition weak when the pileup is high.

TK and PF recoil definitions, seen as mathematical functions, are represented by the identity. The features of the recoil are \( x = \{ h^{TK(PF)}_x, h^{TK(PF)}_y \} \), computed starting from the list of PF candidates, and the output \( y \) is equal to \( x \). TK recoil is measuring only charged tracks and ignore the neutrals. This leads to a measurements of the magnitude that is on average smaller, at least in the high \( p_T \) regime. For this reason, the first basic correction which can be applied is to multiply it by an overall correction factor, in order to measure the correct magnitude on average. In this case, an overall correction factor is roughly 1.8, and this procedure will be done for the sake of comparison.

Usually the systematic uncertainty due to the recoil mismodeling is smaller for TK recoil than for PF recoil, since it does not depend on the simulation of the pileup vertices and on the reconstruction of neutral particles. The latter is usually more difficult than for the charged tracks. This data/simulation discrepancy usually leads to biases of the order of 100 MeV, as seen in the next chapter in section 6.4.2, which is far above the target precision for this measurement. For this reason a specific calibration of the simulation is needed in order to correct for this bias and leave a systematic uncertainty smaller than the target precision.

In this thesis the TK and PF recoil are defined in the following way. First, the lepton(s) identified as the decay product(s) of the W(Z) are removed from the list of the particle flow candidates. Second, all charged tracks with \(|\eta| > 2.4\). The reason of this cut is the following. The absolute position and orientation of the tracker with respect to the frame defined by the beam-axis and the center of the luminous region might not be perfectly reproduced in the simulation. Such discrepancy appears as modification in the tracking efficiency for particles at the borders of the acceptance. In the left plot in figure 5.2, the distribution of the \( \eta \) and \( \phi \) for reconstructed charged particles in Z data is shown. The cut in pseudorapidity removes the problematic region at the edge of the acceptance. A similar problem exists for the calorimeters, but it has been found to be less important and no cut has been applied. Then, the transverse momenta of the remaining PF candidates are summed up to compute the PF recoil. Furthermore, charged tracks coming from the PV are selected, according to an elliptical cut in the 2D plane \( \Delta_{xy} vs \Delta_z \). These variables are the transverse and longitudinal impact parameters. The cut is defined looking at simulated events with zero pileup and defining an elliptical region that contains roughly 90% of the tracks. The right plot in figure 5.2 shows this distribution. The chosen parameters are \( 300 \mu m \) for the \( \Delta_{xy} \) axis and \( 500 \mu m \) for the \( \Delta_z \) axis. This should remove fake and badly reconstructed tracks and guarantee a slightly better resolution on TK recoil. It is important to keep in mind that a change on these parameters does not affect significantly the definitions of TK and PF recoil.

Both these definitions are not satisfactory because they are not exploiting all the information collected by the detector. For instance the angular spread of the particles or the number of recon-
5.1 Defining the recoil

5.1.1 Recoil definition

While the recoil is a 2D vector, its definition is not well established. Let us define the recoil as the difference between the experimental measurement and the generator level recoil:

$$\Delta \vec{h} = \vec{h}_{\text{meas}} - \vec{h}_{\text{true}}$$

First of all, it is possible to look at the 2D distribution of this vector. This 2D distribution depends on the chosen frame in which the components are computed. The frame can be fixed or changing on an event-by-event basis. In figure 5.3 the distribution of the resolution of the TK recoil in W simulated events is shown: in the left plot it is plotted in the fixed basis x-y of the detector, whereas in the right in the basis parallel and perpendicular to the true recoil. The resolution in the first case is almost spherical, except for small deviations due to the small tracker displacement with respect to the beam axis, whereas in the second one it is not spherical at all.

The first plot can give only a qualitative idea of the effect. To get a more quantitative and concrete idea it is possible to compute the effect on the $M_T^2$ distribution with the following reasoning. As shown previously, $M_T^2$ can be expanded in powers of $h/p_l$. In the limit of small $h/p_l$, $M_T^2 \approx 4p_l^2 + 4p_l \cdot \vec{h}$, where $p_l$ stands for the lepton transverse momentum. In the same limit the lepton direction is almost uncorrelated with the recoil one. The corresponding $M_T^2$ resolution

![Image of plots showing 2D distributions of η (y axis) versus φ (x axis) for reconstructed charged pions in Z+ events in the data. The cut |η| < 2.4 has been imposed in order to define well the acceptance in the data and in the simulation. This makes the analysis insensitive to possible discrepancies of the tracker position between the data and the simulation. Right: Distribution of ∆xy versus ∆z for charged tracks in Z+ events with no simulated pile-up. ∆xy and ∆z are defined as the distances on the transverse plane and along the beam axis, respectively, between the PV and the point of closest approach of the track to the beamline.](image-url)

Figure 5.2: Left: 2D distribution of η (y axis) versus φ (x axis) for reconstructed charged pions in Z+ events in the data. The cut |η| < 2.4 has been imposed in order to define well the acceptance in the data and in the simulation. This makes the analysis insensitive to possible discrepancies of the tracker position between the data and the simulation. Right: Distribution of ∆xy versus ∆z for charged tracks in Z+ events with no simulated pile-up. ∆xy and ∆z are defined as the distances on the transverse plane and along the beam axis, respectively, between the PV and the point of closest approach of the track to the beamline. structured tracks might bring additional information that is not exploited in these two definitions of the recoil.

5.1.2 Recoil resolution

The resolution of the recoil is not a well defined concept, because it is a 2D vector whose distribution can be arbitrarily complex and not describable with few parameters. Let us define the resolution as:

$$\Delta \vec{h} = \vec{h}_{\text{meas}} - \vec{h}_{\text{true}}$$

where $\vec{h}_{\text{meas}}$ is the experimental measurement of the recoil and $\vec{h}_{\text{true}}$ is the generator level recoil.

First of all, it is possible to look at the 2D distribution of this vector. This 2D distribution depends on the chosen frame in which the components are computed. The frame can be fixed or changing on an event-by-event basis. In figure 5.3 the distribution of the resolution of the TK recoil in W simulated events is shown: in the left plot it is plotted in the fixed basis x-y of the detector, whereas in the right in the basis parallel and perpendicular to the true recoil. The resolution in the first case is almost spherical, except for small deviations due to the small tracker displacement with respect to the beam axis, whereas in the second one it is not spherical at all.

The first plot can give only a qualitative idea of the effect. To get a more quantitative and concrete idea it is possible to compute the effect on the $M_T^2$ distribution with the following reasoning. As shown previously, $M_T^2$ can be expanded in powers of $h/p_l$. In the limit of small $h/p_l$, $M_T^2 \approx 4p_l^2 + 4p_l \cdot \vec{h}$, where $p_l$ stands for the lepton transverse momentum. In the same limit the lepton direction is almost uncorrelated with the recoil one. The corresponding $M_T^2$ resolution

63
Chapter 5. Towards a better recoil definition

![Recoil TK resolution in the CMS frame](image1)

![Recoil TK resolution in the true boson frame](image2)

**Figure 5.3:** The two plots show the distribution of the resolution of TK recoil in simulated Z+ events, for two different reference frames: the fixed x and y axis of the detector (left) and the parallel and perpendicular axis with respect to the true recoil direction (right). The shapes of the two distributions are extremely different, showing an almost spherical distribution in the first case, and a distribution peaked at 7 GeV in the parallel direction in the second case.

is proportional to the recoil resolution in the muon direction, that is a random direction on an event-by-event basis, as seen from the "recoil point of view". Thus to estimate this resolution is enough to consider the spread of the distribution on one random direction from the "recoil point of view", even if it is not the muon one. In doing so, picking the RMS along the x direction is a simple but good choice, which leads to a spread of about 7 GeV. The expected RMS on $M_2^T$ is $\Delta M_2^T \simeq 4\vec{p}_l \cdot \Delta \vec{h} \simeq 4 \cdot 40 \cdot 7 \text{ GeV}^2 \simeq 1100 \text{ GeV}^2$, where the computation is done considering all the leptons with their maximum transverse momentum, that are those in the peak of the distribution. This is comparable or larger than the natural width of the $M_2^T$ distribution, and makes the peak much broader when passing to reconstructed variables, as observed in the previous studies.

As mentioned, correcting the scale of the TK recoil for the average response is a basic correction which can be applied. This means bringing the mean of the distribution $|\vec{h}_{\text{meas}}|/|\vec{h}_{\text{true}}|$ close to one. However, this does not automatically imply that the resulting resolution is better. To explain this point, figure 5.4 shows a sketch of the true, TK, and scaled TK recoils, in a situation in which after the correction of the scale the resolution does not show an obvious improvement in magnitude.

The distribution of $|\Delta \vec{h}|$ gives another good qualitative idea of the resolution. It does not depend on the reference frame in which the resolution is computed. Additionally, it coincides with the missing transverse momentum, in case the boson decay products were perfectly measured (for instance in the Z case). To rephrase, it is the amount of missing transverse momentum measured...
5.1 Defining the recoil

Figure 5.4: Sketch of a situation in which the angle between the reconstructed TK recoil and the true one (called MC) is large enough so that scaling it in order to have the correct magnitude (STK) is not an obvious improvement. In fact the length of the dashed red line is comparable or even larger than the dashed green line.

because of the poor measurement of the recoil. In fact:

$$\left| \Delta \vec{h} \right| = \left| \vec{h}_{\text{true}} - \vec{h}_{\text{meas}} \right| = \left| -p_T^{\text{boson}} - \vec{h}_{\text{meas}} \right| = \Delta E_T^{\text{miss}}$$

where $\Delta E_T^{\text{miss}}$ takes into account the neutrino in the $W$ events, and it coincides with $E_T^{\text{miss}}$ in the $Z$ case.

The left plot in figure 5.5 shows the distribution of this quantity in simulated $Z+$ events, for various recoil definitions, obtained by scaling the TK magnitude by some constant factors (0, 0.5, 1, 1.6, and 3). Clearly scaling by a factor three (purple line) results in a worse resolution. However it is not trivial to understand which definition is better, since the shape of the distribution of the $\left| \Delta \vec{h} \right|$ changes, in the mean, the peak and the tail.

Additionally, another estimator of the effect of the recoil resolution is the $M_T^2$ resolution, which tells how the transverse mass is affected by the recoil resolution. It is defined as:

$$\Delta M_T^2 = M_{T,\text{meas}}^2 - M_{T,\text{true}}^2$$

and represents the smearing that is applied to the $M_T^2$ distribution due to the imperfect measurement of the recoil. However, guessing the effect on the uncertainty on the $M_W$ fit is difficult.

The right plot in figure 5.5 shows the distribution of $\Delta M_T^2$ for the same recoil definitions used in the left plot. Again it is clear that scaling by a factor three worsens the resolution. Furthermore scaling by 0.5 or 1 results in a slightly narrower peak than using the 0 factor scale, which means no recoil and for which $M_T \equiv p_T$, and the 1.6 scale factor.

The take-home-message of this section is the following: the recoil resolution is a complicated concept, which can be estimated in several ways. The various ways are somehow complementary, showing different aspects of the same concept. However, there is no simple way to estimate the goodness of a recoil definition by looking at the resolution plots. The performance of the recoil definition should eventually be quantified in terms of uncertainties on the $M_W$ fit.
Towards a better recoil definition

5.2 Improving the recoil definition

As previously mentioned, more features of the recoil, related for instance to its geometrical shape or content in terms of charged and neutral particles, might bring additional information. However, there is no unique way to combine this information, i.e. to translate the number of reconstructed tracks into a GeV quantity. Therefore, a way to solve this problem is to combine all the features of the recoil $x$, the input variables of the definition using a multivariate analysis (MVA) algorithm. Particularly, this is a regression problem: we want to predict the value of the true recoil $p_T$ (target), starting from the experimentally measured features of the recoil (inputs).

In the following, while formulating the problem in terms of a regression problem, the standard notation for the regression will be used: $x$ for the independent variable and $y$ for the dependent variable. The independent variables $x$ are the features of the recoil. The dependent variable $y$ is the true $p_T$ of the recoil. These variables have here a different meaning than the one previously explained in section 5.1.

Let us re-formulate the problem in a more schematic way:

- There are some quantities computed on an event-by-event basis, called $x$, and for every event, meaning for every possible set of input variables, we want to predict the true value of the recoil $y$. Let us consider the probability density function $p(y|x)$ and fix a value of $x$. In the ideal case, where the chosen variables identify univocally the recoil value, this pdf is a delta function. In the opposite case in which the variables are completely uncorrelated from the recoil, the pdf results to be a flat distribution. In the normal case, this pdf is a distribution which constrains $y$ in a certain range, usually smaller than the total integrated range of the variable, with values that are more likely than others. This reflects the fact there are many events with the same $x$ values, though corresponding to different $y$. When $p(y|x)$ is known, all the information about the recoil contained in this set of input variables
5.2 Improving the recoil definition

- Assuming perfect knowledge of $p(y|x)$, it is not obvious which value to assign to the recoil. The most probable value, the mean value, the median of $p(y|x)$ are all reasonable choices, and many more can be thought of. Which choice is better is something that should be evaluated in terms of the final goal of the analysis.

- The problem is then translated into a different one: predict $p(y|x)$ for each value of the input variables $x$. Subsequently, different choices of the event-by-event value of the recoil are formulated and evaluated in terms of the goal of the analysis.

The first goal is achieved using the so called semi-parametric regression, which has been implemented on a deep neural network (DNN) architecture.

5.2.1 Semi-parametric regression

To understand what a semi-parametric regression is, we will proceed by steps, extending the concept of a least squares fit, and keeping in mind the actual goal previously illustrated.

To start with, let us recap the least squares fit concepts. Given some data $(x, y)$, a functional form for $y(x) = f(x; a)$ is assumed, where $a$ is a set of parameters which identifies the function. As the pdf $p(y|x)$ is not a delta function in the general case, the previous functional form links the mean of $p(y|x)$ with the $x$ value itself. For instance, in the case of a linear relation between $x$ and $y$, and assuming 1D values for both $x$ and $y$, $a$ corresponds to the slope and the intercept of the straight line, which for each value of $x$ gives the mean of the possible $y$ values. The fit part consists of finding the best value for $a$, and it is achieved by minimizing a certain function of this parameters, namely the sum of the squares of the residuals between the predicted value and the true one:

$$X^2 = \sum_i (y_i - f(x_i; a))^2$$

Once the parameters $a$ are fixed, the function $f(x; a)$ can predict the value of $y$ even for values of $x$ which are not part of the initial data-set.

Going to the machine learning (ML) language, the fitting part is called training, and the second one is called prediction. The function which is minimized in the training phase is the so called loss function.

The previous least square fit is a regression of the mean of $p(y|x)$. A semi-parametric regression is instead a regression of the full pdf $p(y|x)$. It is thus a suitable technique for the problem under study: predicting $p(y|x)$ is indeed the goal of the problem.

To understand the semi-parametric regression concept it is useful to see the previous fit as a semi-parametric regression. In fact the assumption behind the minimization of $X^2$ is that the pdf $p(y|x)$, which is not a delta function, is Gaussian with standard deviation independent on $x$. The proof follows. The functional form $y(x) = f(x; a)$ is actually relating the mean of the
Gaussian distribution of \( p(y|x) \) with the value of \( x \). In this case, computing the Likelihood

\[
L = \prod_i \text{Gaussian}(y_i; \mu = f(x_i; a), \sigma = \text{const})
\]

and passing to the minus logarithm:

\[
-\log(L) = \sum_i \left[ \log(\sqrt{2\pi\sigma}) + \frac{(y_i - f(x_i; a))^2}{\sigma^2} \right]
\]

As \( \sigma \) is constant, the first term and the denominator of the second term can be removed, without changing the result of the minimization, leading to the previous formula for \( \chi^2 \). This loss function is usually called mean square error (MSE). Now, the pdf \( p(y|x) \) can be predicted starting from its parameters: mean regressed and sigma fixed.

Figure 5.6: Left: 2D distribution of the variables \( x, y \): \( y \) is Gaussian distributed, with mean and sigma dependent on the \( x \) variable. The dependence of the mean and sigma on the \( x \) variable is shown in the right plot.

Let us complicate a bit the situation, and suppose that \( p(y|x) \) is Gaussian distributed, but with both \( \mu \) and \( \sigma \) varying as a function of \( x \). This example is illustrated in figure 5.6: in the left plot the 2D distribution of \( x, y \) is shown; the red superimposed bars show the distribution \( p(y|x) \) for various values of \( x \), indicating one standard deviation (red box) and three standard deviations (red lines). Now we have \( \mu = f(x_i; a) \) and \( \sigma = g(x_i; b) \) where \( f \) and \( g \) may be different functional forms. In this example they are shown in the right plot: their functional form is chosen to be a
5.2 Improving the recoil definition

second degree polynomial, with different parameters for the mean and for the sigma. Thus:

\[ X^2 = -\log(L) = \sum_i \left[ \log(\sqrt{2\pi}g(x;b)) + \frac{(y_i - f(x_i; a))^2}{g(x;b)^2} \right] \]

which is thus minimized with respect to both \(a\) and \(b\). It is important to notice that the distribution \(p(y|x)\) should be properly normalized: in this latter case, without normalization, the loss function would become:

\[ X^2 = \sum_i \frac{(y_i - f(x_i; a))^2}{g(x;b)^2} \]

and during its minimization the parameter will change in order to have \(g\) as large as possible which brings \(X^2\) close to zero, which is not the correct result. As before the semi-parametric regression can predict the pdf \(p(y|x)\) for each value of \(x\) through the prediction of the parameters which determines the pdf \(p(y|x)\). In the end, a semi-parametric regression is a regression of the parameters of a functional form which describes the distribution of \(p(y|x)\).

It is noteworthy that the least square method can be used even if the distribution \(p(y|x)\) is not Gaussian with sigma independent on \(x\). In this case, if the correct functional form is assumed for the relation \(y(x) = f(x; a)\), the regressed value \(y(x)\) is going to be the mean of the distribution of \(p(y|x)\). However, this is not a semi-parametric regression, because the distribution of \(p(y|x)\) cannot be predicted, but only its mean.

There are thus two functions that are chosen. The first one is the pdf functional form, needed to fit the probability density function \(p(y|x)\), that in previous cases was assumed to be Gaussian. The second one is the functional form of the parameters which identify \(p(y|x)\) as a function of \(x\): in the previous case shown in figure 5.6 this relation was a second order polynomial both for the mean and for the sigma, but with different parameters for the two cases.

Both functional forms cannot be, in general derived from first principles. Their choice should be guided from the heuristic observation of the behaviour of the data in the training sample and finding appropriate and elastic formulas to regress the parameters. The second functional form, that is the functional form of the parameters of the pdf as a function of \(x\), can be obtained using the deep neural networks.

5.2.2 Deep neural networks

Deep neural networks (DNN) are mathematical functions which depend on a arbitrary large number of parameters, and can be built starting from small blocks called neurons. There are many kind of architectures, such as convolutional, recursive, LSTM, etc., but the following will focus on dense networks, as seen in the left plot of figure 5.7, which have been used in the course of this work.

The idea behind neural networks is to build a sort of universal non-linear function starting from blocks which perform linear operations, which are called neurons, as seen in the right plot of
Figure 5.7: Left: example of a DNN with four layers. All the connections between the neurons are shown. Right: sketch of the workflow of a neuron. The inputs $x_i = 1, \ldots, N$ are multiplied by the weights $w_i$ and then summed up with a bias $b$. An activation function is then applied on the result of the previous operation, producing the output of the neuron.

Each neuron is characterized by some parameters and an activation function and perform the following operation:

$$f\left(\sum_i w_i x_i + b\right)$$

where $i$ runs over the inputs of the neuron, $w_i$ are the weights of the neuron, $b$ is the bias of the neuron, and $f$ is the activation function. The activation function are usually some functions that show a sort of threshold, such as the Sigmoid, the inverse tangent, or it might be even simply the identity function. The activation functions are chosen "a priori" and are part of the architecture, i.e. they are the same function for all neurons in a layer, while the weights $w$ and the bias $b$ depend on the neuron. The weights and the biases are the free parameters of the function: their values are chosen during the training phase, in order to minimize a certain loss function.

Neurons are organized in several layers: in a dense network, often called fully connected network, the output of every neuron is connected as input to every neuron of the next layer. There are two important reasons why such architecture are powerful. Firstly, it can be demonstrated that, provided a certain number of neurons and layers, a DNN can approximate every function. Secondly, the function defined by the DNN is a recursive composition of the various functions of the various blocks. Its derivatives with respect to the parameters can be fast computed using the chain rule. In fact, they result to be the product of the derivatives of each neuron with respect to its parameters, making them easy to compute. One subtlety is the fact that all the neurons in the same layer are equal, as they receive the same inputs and send the output to the same neurons, therefore, there is no apparent reason why the weights and biases of these neurons should converge to different values. The reason why this happens is that the weights and biases are initialized randomly: this breaks the degeneracy from the very beginning, making the weights and biases converge to different values.

In the following case the loss function is chosen to be the negative-log-likelihood of the fit as shown in the previous paragraph, extending the reasoning presented there. Schematically the loss function computed on a certain dataset $(x, y)$ is a function of $y$ and $\text{DNN}(x; a)$, where $a$ stands for the free parameters of the network. To recap, $\text{DNN}(x; a)$ gives as output the parameters which
describe the pdf $p(y|x)$. So the loss function depends on $a$ only through the evaluation of the DNN on $x$. As it is computed many times during the minimization procedure, it is important that the loss function is explicitly written in terms of the parameters which identify the pdf, which in the previous case shown in figure 5.6 were the mean and the sigma of the Gaussian functions. This implies that a close expression for the normalization of $p(y|x)$ is needed, as in the Gaussian case.

For its flexibility and capability of describing every function, the DNN architecture has been chosen.

A more complete review can be found in Ref. [47], whereas an historical overview concerning the evolution of these techniques can be found in Ref. [48]. The Keras library [49] has been used, as it provides a Python wrapper of the TensorFlow library [50]. The complete list of the architecture and training settings which have been used for this regression is explained in appendix B.

### 5.2.3 A semi-parametric regression for the recoil

A few important considerations should be made while applying the semi-parametric regression technique to the recoil problem. The semi-parametric regression is a technique which let us predict the parameters that describe the probability density function of the true recoil, at a fixed value of the input variables. This has many advantages. Firstly, $p(y|x)$ contains all the information that is present in the input variables regarding the recoil. Secondly, given the pdf of the recoil on an event by event basis, it is possible to pick both an estimator of the true recoil, the most probable value of the distribution for instance, and an estimator of the uncertainty on this estimate, that can be the RMS of the pdf distribution for example. There is no preferred a priori rule to define the best estimators: this should be eventually tested by looking at the best performance in terms of systematic uncertainty on the goal of the analysis, i.e. $M_W$.

An important point is the choice of the training sample. In fact, the prediction made by the regression depends on the data-set on which the network has been trained. In the ideal case where the chosen functional form describes perfectly the pdf, the regression returns the $p(h|x)$ as learnt from the training sample, for each set of $x$, where $h$ is the value of the true recoil, which is equal and opposite to the $W-p_T$. It is notable that the regression depends on the $W-p_T$ spectrum in the training sample, in fact:

$$p(h|x) = \frac{p(x|h)p(h)}{\int dp(x|h)p(h)}$$

thus it is important to train on a sample which is as close as possible to the data. Particularly, the training has been performed on the $W$ MC sample, that shows a $W-p_T$ spectrum which is close to the true one. Although the discrepancy in the $W-p_T$ spectrum between data and MC is important in terms of systematic uncertainty on the measurement, there is no bias which depends on the choice of the training sample, as long as the recoil definition which is applied on data is the same as the one applied on the simulation. The only problem which could arise is
the non-optimality of the recoil definition, in case we are using an incorrect value, on average, for the recoil on an even-by-event basis. We expect that data/MC discrepancies in the $W-p_T$ spectrum are small with respect to the experimental resolution of the recoil: the effect of the $W-p_T$ discrepancy between training sample and data will be thus negligible.

A first subtlety is the fact that the recoil (the $y$ variable) is a vector and therefore has two components. Therefore a 2D semi-parametric regression is needed. Parametrizing the 2D distribution of the true recoil for each value of the input space and using just few parameters is non-trivial. A way to overcome this problem is to use as target two variables which describe entirely the recoil (for instance magnitude and angle, $x$ and $y$ components, etc.) and train two different semi-parametric regression. The smaller the correlation between the two variables, the better is the result of this approach.

A second issue is that it is important to have variables that are Gaussian-like distributed, meaning that they have a peak, as sharp as possible, with tails, possibly non-Gaussian. A distribution like this can be easily fitted with crystal-balls functions. Furthermore, from a distribution like this it is easier to choose an estimator of the recoil on an event-by-event basis: the $x$-value of the peak for instance make sense, and if the tails are symmetric, it coincides with the mean. If the distribution had more peaks for instance, it would be much more difficult to define the best estimator. In addition, the target of the regression should be a variable that does not change significantly on an event by event basis: this can make the training faster and more stable.

For these reasons, instead of targeting the two components of the recoil, it is possible to consider the correction that must be applied to a certain reconstructed recoil (TK or PF for instance, as explained in 5.1.1), and use it as target for the regression. In fact, whereas the recoil is something that moves from zero to several tens of GeV, the coefficients that describe the correction could be something of order one for all the values of the input space. Additionally, using correction coefficients better exploits the information contained in the TK or PF recoil, since they are used as starting points of the new recoil.

Furthermore, it is useful to recap the idea of the recoil definition as a mathematical function of the inputs. This new definition is fixed by the choice of the input variables, the targets, the DNN model, the training sample, the training itself, and the procedure to pick the estimator on an event-by-event basis. All of these choices are arbitrary, but it is important to stress that there is no intrinsic systematic uncertainty directly related to this choice: a bad choice of these parameters results in a sub-optimal recoil definition. Recalling that a good recoil definition is useful for the mitigation power of the $W-p_T$ uncertainty, being sub-optimal means that this power has not improved.

To conclude this set of considerations, an important point of the recoil definition is the systematic uncertainty due to a mis-modeling of the inputs: with a new definition based on the semi-parametric regression it is expected to be larger than for TK, as more information is used.
5.2 Improving the recoil definition

5.2.4 Possible parametrizations of the correction

As previously mentioned, it is useful to use correction coefficients to traditional recoil definitions as the y variables, i.e. the targets of the regression. The correction can be parametrized in several ways. Two of them, taken into account in the course of this thesis, are $c_{1,2}$ and $e_{1,2}$. Their description is sketched in figure 5.8. $c_{1,2}$ parametrize the correction in terms of parallel and perpendicular component with respect to the measured recoil. $e_{1,2}$ parametrize instead the correction with a scale factor of the magnitude, and the azimuthal angle by which the measured recoil is to be rotated.

\[
\begin{align*}
    c_1 &= h_{MC}^\parallel / h_{tk} \\
    c_2 &= h_{MC}^\perp / h_{tk} \\
    e_1 &= h_{MC} / h_{tk} \\
    e_2 &= \Delta \phi
\end{align*}
\]

Figure 5.8: As the correction that must be applied to the reconstructed recoil ($\vec{h}_{TK}$) in order to get the true one ($\vec{h}_{MC}$) is a 2D vector, it can be parametrized in several ways. Two possibilities studied in the course of this thesis are shown: $c_{1,2}$ (left) and $e_{1,2}$ (right).

It is useful to look at the distribution of these coefficients in order to understand which to choose. A comparison between some of these coefficients is visible in figure 5.9, and shows that the distribution of $c_3$, $1/c_1$, $e_1$, and $\ln(e_1)$. The first three present some complex features, such as broad and significantly asymmetric peaks, double peaks, and points were the pdf is identically zero, whereas the last one shows a nicer distribution with only one peak with two almost symmetric and smooth tails.

5.2.5 Evaluating the goodness of a definition

There is no single, simple check in order to validate the choice of a recoil definition, being it a traditional definition like TK or a more complex one obtained with a semi-parametric regression. It is thus useful to split the problem in several steps. First of all, in case of an algorithm like the semi-parametric regression that relies on a minimization procedure during the training phase, the convergence of this algorithm is checked with several control plots. Secondly, it is important to look at the resolution of physical observables, like the recoil magnitude and angle, and the
transverse mass. It gives an idea of the possible results in terms of the uncertainty on the $M_W$ fit. Eventually, the most important quantity to evaluate the goodness of a definition is the event-by-event correlation between the lepton and the recoil, which is what makes the transverse mass independent on the $W$-$p_T$ spectrum. No way to quantify this correlation with just one number has been found. This has been better quantified using a fitting technique, in the framework of the systematic uncertainties, that will be discussed in detail in the next chapter.

5.3 $e_{1,2}$ regression

The regression with the coefficients $e_{1,2}$ is the best regression obtained in the course of this thesis, and it is thus described in the details in this section. The results obtained with this recoil definition will be also analyzed in terms of systematic uncertainties in the next chapter.

5.3.1 $e_{1,2}$ distributions

To recap, the definitions of the corrections to the a certain definition of the recoil which were chosen are $e_1$ and $e_2$. They correspond to the correction to the magnitude of the measured recoil ($e_1$) and the correction in the azimuthal angle ($e_2$).

In order to describe the features of $e_{1,2}$, it is important to fix the starting point of these coefficients. It has been decided to start from TK instead of PF because of two reasons. The first one
is that TK slightly better measures the recoil direction, that is expected to be more difficult to correct compared to the magnitude. In fact, the presence of an additional vector is required in order to remove the degeneracy left-right. In fact, as long as no additional vector is added to the inputs, the distribution of $e_2$ is symmetric. A plot of $e_2$ in the two cases of TK and PF is shown in figure 5.10. The distribution in case of TK recoil is narrower than PF. The second reason is that discrepancies between data and MC are expected to be smaller for TK than for PF, i.e. modeling in the simulation TK is easier than PF. For this reason, the systematic uncertainty due to a mis-modeling of the recoil is expected to grow when adding information concerning neutral particles.

![Figure 5.10: $e_2$ distributions for TK and PF recoil, as computed in Z simulated events. The former has a narrower distribution than the latter, meaning that on average the angle is better measured. Since the angle is expected to be more complicated to be corrected, it has been chosen to start from TK.](image)

From now on, $e_1$ and $e_2$ will refer, unless explicitly written, to the correction that are applied to the TK recoil in order to transform it into the true one.

**Magnitude correction $e_1$** It is a variable positive definite, and presents a long tail at high values due to events where the measured recoil is small, because some tracks are lost for the acceptance or because neutrals took most of the momentum of that events. As previously shown in figure 5.9, the distribution of the log($e_1$) is more suitable for a semi-parametric regression. In figure 5.11 the distribution of log($e_1$) in bins of the magnitude of TK recoil is shown. The evolution of the curves from small values (black) to large value (green) highlights the behaviour of the recoil between two different regimes. In the case of large TK recoil, the recoil is in the jet regime, and log($e_1$) peaks at around 0.4, corresponding to $e_1 \approx 1.5$, that is the value to correct for the absence of the neutral particles in the recoil. At small TK recoil the distribution is instead much broader, with a broader peak and longer tails at larger values. This is due to the fact that $e_1$ is the ratio of the true recoil by a small number and that the TK recoil shows very little correlation with the true one.
Figure 5.11: Distribution of the log($e_1$) in bins of the magnitude of the TK recoil. The distribution of this correction varies significantly between the various bins: in the case of small recoil (black curve), it is really uncorrelated with the true one, resulting in a correction with a broad distribution. In case of large reconstructed recoil (green curve), that corresponds to the jet regime, the correction mainly takes into account the lack of the neutral particles.

The functional form chosen to fit the pdf of this coefficient is the Gauss-double-expo-tails (GD). It is a function made of a Gaussian core, matched in a $C^1$ way to two different exponential tails on the two sides. Mathematically:

$$f(y|a(x)) = \frac{1}{N} \begin{cases} 
  e^{\alpha_1^2/2 + \alpha_1 t}, & \text{for } t < -\alpha_1 \\
  x(n-1), & \text{for } -\alpha_1 \neq t \neq \alpha_2 \\
  e^{\alpha_2^2/2 - \alpha_2 t}, & \text{for } t > \alpha_2 
\end{cases}$$

where $t = \frac{y - \mu}{\sigma}$ and $a = \{\mu, \sigma, \alpha_1, \alpha_2\}$ are the parameters of the distribution, and

$$N = N(\sigma, \alpha_1, \alpha_2) = \sqrt{\frac{\pi}{2}} \sigma \left[ \text{erf} \left( \frac{\alpha_2}{2} \right) - \text{erf} \left( \frac{\alpha_1}{2} \right) \right] + \frac{e^{-\alpha_1^2/2}}{\alpha_1} - \frac{e^{-\alpha_2^2/2}}{\alpha_2}$$

is the normalization factor. The function has, in principle, nine free parameters, mean, sigma and normalization for the Gaussian core, decay length, starting point and normalization for each of the two exponential tails. There are four constraints: two for the continuity and two for the derivatives continuity at the matching points. In addition the overall normalization is fixed to one. This leaves the function with only four free parameters: mean ($\mu$) and sigma ($\sigma$) of the Gaussian core, and the two exponential decay lengths ($\alpha_1$ and $\alpha_2$). This function is very elastic in fitting distributions with one, possibly asymmetric peak, with two Gaussian or exponential tails.
Angular correction \( e_2 \) It is defined in the interval \([-\pi, +\pi]\) by construction, and its distribution is continuous in \(-\pi = +\pi\) because applying a correction of \(-\pi\) is equivalent to applying a \(+\pi\) correction. Therefore, a fit function for \( e_2 \) should be at least continuous in this point. However, it is difficult to identify a function with this property, which also respects all the other previous requirements. Therefore this request has not been addressed: anyhow this does not represent a problem, as it is important that the function describes fairly well the distribution, even if it is not correct one.

The chosen function for \( e_2 \) is GD, restricted to the interval \([-\pi, +\pi]\) plus a constant noise distribution (GDC). This adds one free parameter that is the coefficient of the constant term. This constant term is important because the distribution of \( e_2 \) does not go to zero near the boundaries of the interval.

5.3.2 Input variables

The input variables for this regression are chosen in order to characterize the recoil as accurately as possible. Variables concerning the magnitude of the recoil, its composition in terms of particles, and its geometrical shape are computed.

It is of interest to notice that all the input variables which have been used are "high level quantities", features of the recoil computed starting from the reconstructed particle flow candidates.

However, exploiting the power of the neural networks and implementing more complex architecture, it might be possible to use the whole low level information. This could be done by using as input variables the whole list of PF candidates (for instance using a recursive neural network), or a photography of the detector (using then a convolutional neural network). However these architectures require larger training samples and a more sophisticated training phase. For this reason it has been chosen to start with this simpler architecture.

To begin with, let us split the particles in several recoil vectors, obtained by the vectorial sum of the transverse momenta of the considered set:

- charged tracks from the primary vertex, as previously described (TK)
- all the neutral particle flow candidates (NT)
- all the charged tracks that are not coming from the primary vertex (NPV)
- the sum of NT and NPV (NTNPV)
- all the particle flow candidates, as previously described, TK + NTNPV (PF)

It can be noticed that, in case of perfect resolution and acceptance, the NTNPV recoil should correspond to the neutral part of the true recoils. To refer to a certain recoil the following syntax will be used: \( \vec{h}_{TK} \) refers to the vector TK recoil, whereas \( |\vec{h}_{NTNPV}| \) refers to the magnitude of the recoil NTNPV. The same abbreviations, where not followed by recoil, will refer to the set of particles.
The variables used for the regression are now split in those common to both $e_1$ and $e_2$, those used only for $e_1$, and those used only for $e_2$.

Both

- The magnitude $|\vec{h}_{TK}|$ and the azimuthal angle $\phi_{TK}$ of the TK recoil. The angle is important since displacement of the detector produce an asymmetric distribution in the azimuthal angle.

- The number of charged tracks present in TK

- The magnitude of the NTNPV recoil $|\vec{h}_{NTNPV}|$

- The $p_T$ of the track (among TK) with the highest $p_T$ (leading track)

- The $p_T$ of the neutral (among NT) with the highest $p_T$ (leading neutral)

- The invariant mass of TK $m_{TK}$

- The sphericity in the transverse plane$^1$ of TK, NTNPV, and PF

- The number of reconstructed vertices

In order to get a faster convergence of the networks, the input variables should be as Gaussian as possible and of order one. In order to get these conditions satisfied, all the quantities measured in GeV but the $p_T$ of the TK recoil are divided by the $p_T$ of the TK recoil and the logarithm of this ratio is taken. This computation narrows the distribution and produces shorter tails. In addition all the variables are rescaled in order to have the bulk of the distribution between zero and one. As this procedure is done only to help the convergence, the rescaling is done with a linear transformation, whose parameters are chosen arbitrarily.

Only $e_1$ In addition, for what concerns $e_1$, the cosine of the angle in the transverse plane between TK recoil and NTNPV $\cos(\phi_{TK} - \phi_{NTNPV})$ is added. In fact, this angle should distinguish between the jet regime and the soft $p_T$ regime: in the former case the angle should be close to zero, in the latter it should be almost random.

Only $e_2$ For what concerns $e_2$, as long as only the previously mentioned variables are used, the distribution is symmetric around zero. Indeed an additional direction is needed in order to break the degeneracy between positive and negative angles around the TK recoil. The angle that is expected to contain most information is $\phi_{NTNPV} - \phi_{TK}$, since, as already mentioned, $\vec{h}_{NTNPV}$ corresponds to the direction of the neutrals from the PV in case of no pileup.

---

$^1$The sphericity in the transverse plane is defined as the ratio between the vectorial sum and the scalar sum of the transverse momenta of the particles belonging to a certain set. In formula: $\frac{|\sum_i \vec{p}_i|}{\sum_i |\vec{p}_i|}$ where $i$ goes through the various particles.
Two more angular inputs are added: the one between the TK recoil and the leading charged track $\phi_{\text{leading track}} - \phi_{TK}$, and the one between the TK and the PF recoil $\phi_{PF} - \phi_{TK}$. Thanks to the symmetry, switching all the signs of the angles leaves the distribution of $e_2$ unchanged apart for a minus sign. Therefore, in order to exploit this symmetry the previous angles have been modified in the following way:

- $|\phi_{NTNPV} - \phi_{TK}|$
- $(\phi_{\text{leading track}} - \phi_{TK}) \cdot \text{sgn}(\phi_{NTNPV} - \phi_{TK})$
- $(\phi_{PF} - \phi_{TK}) \cdot \text{sgn}(\phi_{NTNPV} - \phi_{TK})$

and $e_2 \cdot \text{sgn}(\phi_{NTNPV} - \phi_{TK})$ is used as target instead of $e_2$. The $e_2$ correction that is therefore applied is multiplied by $\text{sgn}(\phi_{NTNPV} - \phi_{TK})$. This operation exploits the symmetry and helps the convergence of the algorithm.

5.3.3 Picking a recoil estimator from the distribution

The semi-parametric regression gives as output the parameters that describe the distributions of $e_1$ and $e_2$ for each set of input variables. In order to produce a new recoil definition, an estimator of the recoil must be chosen from these distributions. Given this event-by-event estimator the correction is trivially applied in the following way:

$$h_{\text{reg}} = h_{TK} \cdot e_1 \quad \text{and} \quad \phi_{\text{reg}} = \phi_{TK} + e_2 \cdot \text{sgn}(\phi_{NTNPV} - \phi_{TK})$$

The two easier way to choose a value for $e_1$ and $e_2$ from their distributions is taking the mean or the most probable value (MPV) of the distributions. More precisely, two estimators of the magnitude and two estimators of the angle are produced.

As $e_2$ is a circular quantity, the mean of $e_2$ must be carefully computed, meaning that it must not depend on the choice of the angle zero. The proper way to do it is to compute the mean of the distribution of the sine ($\sin\text{mean}$) and of the cosine ($\cos\text{mean}$) of $e_2$, and consider the angle determined as $\text{atan}(\sin\text{mean} / \cos\text{mean})$. This computation is long, and the result is very close to the MPV of the distribution: therefore only the MPV of the $e_2$ distribution is computed and used for the recoil definition.

Eventually two recoil definitions are produced and analyzed: peak\(^2\) (PK) is obtained by using the MPV both for the magnitude and for the angle, whereas mean-peak (MNPK) is obtained using the mean for $e_1$ and the MPV for $e_2$.

5.3.4 Closure plots

Various plots are produced to visually evaluate the convergence of the regression, and to compare the goodness of the various definitions. A more quantitative comparison between the various re-

\(^2\)It is called peak because the MPV corresponds to the peak of the distribution.
Checking the convergence  

First of all, the convergence of the algorithm is carefully checked. This step is not fundamental: even a bad convergence of the algorithm does not imply any bias on the $M_W$ measurement as long as the same definition is applied both on the data and on the simulation. Checking the convergence means assessing the agreement between the true and the predicted distributions of the target. In this phase the recoil definition is not important: only the capability of the regression in predicting the correct target distribution $p(y|x)$ is evaluated. There is no analogous step in case of the simpler definitions TK and PF.

This is done with the so-called sum-of-pdfs plots, figure 5.12. Given a testing sample\footnote{A testing sample is an homogeneous sample to the one on which the training has been performed. It is useful to evaluate the performance of a neural network on a different sample because this might unveil possible over-training. Over-training is what happens when a neural network starts to "learn" the statistical fluctuations of a sample.} two histograms are filled and compared. The first one is filled with the true target variable, computed for each event. The second one is filled in the following way. For each event, the regression is evaluated on the input variables producing the predicted probability density function $p(y|x)$ of the target for that event. Then the histogram is filled with the full probability density function normalized in order to have unitary integral. The two histograms are then compared: in case the description of the target distribution and the convergence of the minimization algorithm are perfect, the two distributions must agree perfectly, meaning that the pdf predicted by the regression is reproducing perfectly the true one. The sum-of-pdfs plots for $\log(\varepsilon_1)$ (left plot) and for $\varepsilon_2$ (right plot) are shown in figure 5.12 in case of $W$ simulated events, where the regression has been trained.

Good agreement can be seen between the two histogram, although the discrepancies, of the order of ten percent, are clearly not due to statistical fluctuations but to the fact that the GD and GDC, the functions which have been chosen for the fit, are not perfectly describing the true distribution in the whole phase space.

It is important to notice that this check is a necessary condition, and not a sufficient one. In fact, this check, which has been performed on the whole sample, should in principle be repeated for each bin of the input space ($x$ variables).

Secondly, this plot is reproduced for the recoil components. Two parametrizations are studied: magnitude and angle, and $x$- and $y$- components of the recoil in figure 5.13 and 5.14. The two plots contain more information with respect to figure 5.12, so, we will proceed by steps. First of all, it is of interest to compare the target distribution and the sum-of-pdfs distribution for magnitude and angle, and $x$- and $y$- components: this are shown with the same colours as previously, blue for the sum-of-pdfs and black for the true target distribution. In the ratio canvas the ratio of the sum of pdfs distribution by the true generator level is shown (it is the reciprocal of the plots in figure 5.12). This comparison shows the same agreement as seen in the previous
plots, and the discrepancies, of the order of few percents, reflect the discrepancies seen in the previous plots. Furthermore, the distribution of the scaled TK recoil (STK) is shown in the same plot (green histogram). This plot aims to make a comparison between the best way to reproduce the $W$-$p_T$ spectrum using the TK recoil and the result of the new method based on the regression. Let us focus on the left plot of figure 5.13, which highlights the main point of the reasoning. STK on average measures correctly the $W$-$p_T$, however, it does not reproduce correctly the integrated $p_T$ distribution, with discrepancies of the order of 20% as seen in the ratio canvas. The sum-of-pdfs method measures reproduce correctly the $W$-$p_T$ spectrum. This opens the possibility of a precise measurement of the $W$-$p_T$ spectrum in the low $p_T$ region, as better explained in appendix A. It is important to keep in mind that this result has been achieved with a new technique, that is a step further the simple measurement of the TK recoil. The right plot of figure 5.13 shows the distribution of the azimuthal angle of the recoil. It shows TK instead of STK, because the scale factor does not affect the angle. As seen in this plot, the true $\phi$ spectrum (black line) is almost flat, because the production is symmetric. A small modulation can be seen, and it reflects the non-flat efficiency in the $\phi$ angle in reconstructing muons. The sum-of-pdf reproduces well the true distribution: again, the discrepancy reflects the imperfect description of $e_2$ shown in the right plot in figure 5.12. The TK recoil (green line) shows an important modulation, of the order of ten percent: it is due to the displacement of the center of the tracker with respect to the beam-axis.

Similar considerations are valid for the two plots in figure 5.14, which show similar plots but for the two Cartesian components of the recoil in the CMS reference frame.
Figure 5.13: Sum-of-pdfs closure plots for the magnitude (left) and angle (right) of the recoil. The agreement between the sum-of-pdfs distributions (blue) and the generator level quantity ones (black) is fairly good, with discrepancies at the level of few percents. Such discrepancies reflect what seen in figure 5.12.

The distribution of STK is also shown in the plot of the magnitude of the recoil, and the distribution of TK is shown in the plot of the angle. These plots highlight the fact that the new method based on the regression is able to reproduce the $W$-p$_T$ spectrum in a better way than using only the TK recoil.

Figure 5.14: Sum-of-pdfs closure plots for the two Cartesian components of the recoil, p$_x$ (left) and p$_y$ (right). The two plots are very similar and show similar aspects to what already seen in figure 5.13.
Resolution on physical quantities  After assessing the goodness of the regression in terms of description of the target distribution and convergence of the training, some qualitative checks of the improvement of the recoil resolution are made. The previous check were in fact performed on integrated distributions. They do not tell anything about the goodness of the variables in terms of uncertainty on the $M_W$ fit. As mentioned, the easiest way to use the information provided by the regression is to pick an estimator of the recoil from the distribution of $p(y|x)$ on an event by event basis\textsuperscript{4}. The two recoil definitions PK and MNPK based on the regression, are obtained after applying a correction on the bare TK recoil, as discussed previously in 5.3.3. These new definitions should bring to a better event-by-event measurement of the recoil. The following plots are also meant as a visual comparison between the different recoil definitions.

First of all, the scale correction factor ($e_1$ ) and the angular correction factor ($e_2$ ) are computed for the new recoil definitions. These residual corrections are compared to the previous one in the two plots in figure 5.15. The two plots contain also the definitions PF and STK. In the left plot the distribution of the log($e_1$) is shown: the value zero means that there is no correction to apply. TK is peaking at a negative value, showing that it is, on average, underestimating the magnitude of the recoil. STK is peaking at zero, meaning that the overall correction factor of 1.8 allows a measurement that is on average correct. MN and PK on the other hand are both peaking close to zero but they are narrower than the STK distribution: the residual correction is on average smaller than the correction that must be applied on TK.

Analogous reasonings are applied for the right plot, which shows the distribution of $e_2$ . This correction is symmetric for positive and negative values and thus its goodness is related to the width of the distribution. For the PK case, the plot shows that the distribution of the correction is narrower, meaning that on average the angle for which to correct PK is smaller than the one to correct TK.

In both the two cases PF shows a worse resolution than the other definitions: for what concerns the scale correction it usually overestimates the magnitude of the recoil, whereas the distribution of the angular correction is broader than the other two definitions.

The left plot in figure 5.16 shows the integrated distribution of the magnitude of the recoil for the various definitions. The distributions are different and in general do not reproduce the true one. This is not an event by event comparison, that is needed for judging the resolution in a proper way, as shown in the right plot. This plot shows the distribution of the recoil resolution $|\Delta h|$ for the various recoil definitions. Multiplying TK by the scale factor worsens the resolution that results to be broader and shifted to larger values. Both PK and MNPK show improvements with respect to the bare TK, although only PK is improving also with respect to STK. PF has the worst resolution among all the compared definitions. However, in all of these cases, it is difficult to evaluate the best definition as the distribution of this variable has features that are difficult to compare, such as the position of the peak and the importance of the tails.

Furthermore, the residual correction in the x and y reference frame is taken into account. These

\textsuperscript{4}This is the easiest because the regression, providing the full pdf $p(y|x)$ can in principle be exploited for a more complicated fit than the template fit to the $M_T^2$ distribution.
two corrections have been called as $\Delta_x$ and $\Delta_y$. Analogous considerations are valid for these two plots (figure 5.17), which are extremely similar by symmetry: in this case, the residual correction to apply in the case of PK is only a bit smaller than the one to apply on STK, whereas the MNPK case looks similar to the TK one.

Eventually the event by event resolution on $M^2_T$ is shown in figure 5.18. This quantity represents the additional smearing that is applied to the $M^2_T$ distribution due to the recoil resolution, and it affects somehow the statistical uncertainty that is expected on the $M_W$ measurement.

It is thus expected a fair reduction of the statistical uncertainty in the case of PK with respect to TK. On the other hand, STK and MNPK are expected to show a similar statistical uncertainty, but larger with respect to TK. PF should result in the largest statistical uncertainty among the variables that have been studied.

The bottom line of all these control checks is that the new recoil definitions obtained with the semi-parametric regression technique are indeed improving the resolution on the magnitude and on the angle. The improvement is instead not manifest in terms of recoil and transverse mass.
5.4 Conclusions

In this chapter the problem of the recoil definition has been formalized and studied in the detail. The fact that the recoil is a 2D object has been stressed and the resolution has been characterized.

Figure 5.17: Comparison of the various definitions in terms of the residual corrections on the x (left plot) and y (right plot) axes, called $\Delta_x$ and $\Delta_y$.

Figure 5.18: $M^2_T$ resolution for the various definitions that have been studied. As the width of this distribution is expected to be related to the statistical uncertainty in the $M_W$ fit, the various definitions are expected to result in similar statistical uncertainties, a part for PF that shows a far larger width than the others.

resolution. As a result, these plots are not enough to judge the various recoil definitions: the performance must be evaluated in terms of systematic uncertainty on the $M_W$ fit.

5.4 Conclusions

In this chapter the problem of the recoil definition has been formalized and studied in the detail. The fact that the recoil is a 2D object has been stressed and the resolution has been characterized.
Chapter 5. Towards a better recoil definition

A new method for improving the definition using machine learning techniques in order to correct the "poor man" definitions (TK and PF) has been developed. Closure checks in order to visually evaluate the convergence of the regression and to compare the various definitions in terms of recoil resolution are made. These checks show some improvements in the recoil, particularly for what concerns the residual corrections in magnitude and angle, that however do not directly reflect in an improvement in the transverse mass resolution.

The next step is to quantify the performance of the different definitions in terms of systematic uncertainties. In fact, the main reason for using the recoil and the transverse mass is to mitigate the uncertainty related to the imperfect knowledge of the $W-p_T$ spectrum; however, a more complex definition, which exploit more information, is more sensible to possible discrepancies between the data and the simulation, which reflect in an additional systematic uncertainty.
Chapter 6

Systematic uncertainties

The correct assessment of the systematic uncertainties is the most crucial part of the analysis. In this chapter the framework in which the systematic uncertainties are treated is shown, stressing its effectiveness in distinguishing the sources of the uncertainty and decomposing them into uncorrelated uncertainties. In this framework, the systematic uncertainty due to the imperfect knowledge of the $W-p_T$ spectrum is analyzed, in order to rank the most effective definitions of the recoil in mitigating this uncertainty. The systematic uncertainty due to an imperfect modeling of the hadronic recoil is treated subsequently. First of all, a method to obtain corrections for the recoil based on a comparison between $Z$ data and MC events is developed, and its effectiveness is estimated. Secondly, the issues in applying these corrections to the $W$ data are illustrated together with possible solutions.

6.1 Factorization of the systematic uncertainties

In order to treat systematic uncertainties that are as uncorrelated as possible, it is important to factorize the different contributions. Since the extraction of $M_W$ is done through a template fit, systematic uncertainties arise from discrepancies between the data and the simulation, which result in templates that differ from data more than just by their statistical fluctuations. Only three variables are needed to compute the transverse mass $M_T$: the magnitude of the lepton transverse momentum, the magnitude of the recoil transverse momentum and their relative angle in the transverse plane. The correlation between the lepton variables and the recoil variables arises from the $W$ decay: at a fixed value of the $W$ kinematics, the distributions of the lepton kinematic variables and of the recoil variables are uncorrelated. The $W$ kinematics is identified with a set of variables $z$, which completely determines it, such as $p_T$, $p_z$, $\phi$ and $M$.

The probability density function of the transverse mass can be factorized in order to show the
dependence on $z$ explicitly:

$$f(M_T) = J(M_T; p_T^l, \Delta \phi_{l,W}, p_T^h, \Delta \phi_{h,W}) \times f(p_T^l, \Delta \phi_{l,W}, p_T^h, \Delta \phi_{h,W})$$

$$= J(M_T; p_T^l, \Delta \phi_{l,W}, p_T^h, \Delta \phi_{h,W}) \times \int f^1(p_T^l, \Delta \phi_{l,W}|z) f^2(p_T^h, \Delta \phi_{h,W}|z) f^3(z) dz$$

where the $f^i$ are the probability density functions of their arguments, $J$ is the Jacobian of the $M_T$ definition, which transforms the kinematic variables of the lepton and the recoil to $M_T$. The relative angle between the lepton and the recoil has been decomposed in $\Delta \phi_{l,W}$ and $\Delta \phi_{h,W}$, which are the relative angles between the lepton and the $W$, and the recoil and the $W$, respectively. This decomposition is useful to separate the kinematics of the lepton from the recoil.

The dependence of $f^3$ on $z$ is now carefully analyzed. Since the distribution is invariant under rotations around the $z$ axis, the angle $\phi$ can be neglected. This is valid only in the assumption that the detector is invariant under $\phi$ rotations, that it is not completely exact, but it is verified at this level of approximation. The independence of $f^3(z)$ on $M$ is crucial, since the assumption behind any calibration of the recoil computed on the $Z$ samples is that this calibration can be ported to the $W$ only if it is independent on the mass of the boson. This means that, at a fixed value of the kinematics of the boson, the distribution of the recoil is independent on the mass of the boson, and thus is the same for both the $Z$ and for the $W$ cases.

The two probability density functions $f_1(p_T^l, \Delta \phi_{l,W}|z)$ and $f_2(p_T^h, \Delta \phi_{h,W}|z)$ are sort of transfer functions: they describe how to pass from the generator level boson to the reconstructed level of the objects, and thus contain the effect of the boson decay and of the hadronization of the recoil, as well as the acceptance and the resolution of the detector. In this framework $f_1(p_T^l, \Delta \phi_{l,W}|z)$ and $f_2(p_T^h, \Delta \phi_{h,W}|z)$ at a fixed value of $z$ are uncorrelated. The effect of discrepancies in the recoil can be studied without considering the lepton, assuming that the lepton description is perfect.

Recalling the notation used in the previous chapter, the recoil definition can be treated in an abstract way as a function, whose inputs $x$ are features of the recoil, quantities experimentally defined and measured. The output are instead the kinematic variables that describe the recoil $y$: they are the magnitude and the azimuthal angle, or any new re-parametrization of these two.

$$y = R(x)$$

So the recoil definition concentrates into two variables all the information about the recoil relevant for $M_T$. Thus, the probability density functions are related by:

$$f(y) = \frac{dx}{dR} f(x)$$

where $dx/dR$ is the Jacobian of the transformation that brings $x$ to $y$. Therefore:

$$f(y|z) = \frac{dx}{dR} f(x|z).$$
Once the recoil definition is fixed, all the discrepancies in the recoil distribution arise from discrepancies in the features of the recoil at a fixed value of $W-p_T$.

In this framework, any new parametrization of the recoil kinematics is acceptable as $y$ variables. As all these reasonings are done at a fixed value of the $W$ kinematics, i.e. of the $z$ variables, $p_T^h, \Delta \phi_{h,W}$ are equivalent to $p_T^h, \phi_h$ apart for a change of variables. Therefore, in the following, the pair of variables $p_T^h, \Delta \phi_{h,W}$ will be chosen as the $y$ variables.

### 6.2 Fit description

In order to evaluate the various systematic uncertainties, a fit is performed to assess the bias due to a certain discrepancy between templates and data. In order to isolate the effect of a certain discrepancy, the fit is performed using the same sample both for templates and pseudo-data. Templates are obtained from the sample by injecting a discrepancy in one of the probability density functions mentioned before. This is done by reweighting or changing the values of a certain variable. Pseudo-data is just the pure data-sample as it is, though the problem is symmetric and the discrepancy can be injected in the pseudo-data as well. Other possible discrepancies, that are not of interest, like the lepton calibration, cancel out: the bias obtained by the fit assesses directly the systematic uncertainty. The statistical fluctuations cancel out too, but not completely: a certain degree of fluctuation, much smaller than what is expected in case of independent samples, is still present, and it is due to the finite statistic of the sample. It will be better discussed in the following.

The templates for the various masses are obtained through a Breit-Wigner reweighting: in fact, thanks to the relatively large width of the $W$ boson, copies of the sample with different masses can be created by reweighting the sample with the ratio of the Breit-Wigner distributions at the new mass by the generated mass in the simulation. This procedure is correct only to reweight the mass in a range smaller or comparable to the width of the $W$ boson. More details about this procedure can be found in appendix C.

The fit is performed binning the data and making a $\chi^2$ scan,

$$\chi^2 = \sum_i \frac{(n^T_i - n^D_i)^2}{n^T_i}$$

where the index $i$ loops over the bins, and $T$ and $D$ refer to templates and pseudo-data respectively. The assumption behind the $\chi^2$ minimization is that the uncertainty in each bin is Gaussian: this is true as long as the number of entries in each bin is large enough to approximate a Poisson distribution with a Gaussian distribution. The points are subsequently fitted with a parabola. The position of the minimum gives the bias, that is the systematic uncertainty. The width of the parabola, that is the width of the range whose boundaries are the points where the $\chi^2$ increases by one unit with respect to the minimum, gives the statistical uncertainty.

It has been verified that the choice of the number of bins is not important as long as a relatively
large number of entries is present in each bin: this makes the approximation of Gaussian errors valid in each bin. The number of bins is usually fixed at 50. In addition, to properly compare different variables, the fit range has been chosen using quantiles: in this way the same fraction of events is present in the histograms filled with different $M^2_T$ definitions. The range is usually fixed considering the events between the quantile $q = 0.25$ and $q = 0.9$. The fit result depends in an important way on the fit range: its choice will require a careful and optimal definition before the final measurement. For the purpose of this thesis it has been chosen in this way in order to have stable, reasonable, and comparable results.

In addition, since the events close to the peak of $M^2_T$ are the most sensitive to $M_W$, and they are those in which the lepton decay happens on the transverse plane in the $W$ rest frame, a cut on the lepton $p_T$ is applied. This cut is $38 \text{ GeV} < p_T < 60 \text{ GeV}$ for the $Z$ events, and the boundaries are rescaled by the ratio $M_W/M_Z$ for the $W$ events. In addition, a cut on the recoil is added in order to restrict the contribution from high $p_T$ bosons. This cut is chosen to be $h_{TK} < 20 \text{ GeV}$ for $Z$ events, and it is rescaled as for the lepton $p_T$ boundaries. There is no study behind the choice of these boundaries and cuts for the fit, other than being guided by the previous reasonings. However, the purpose of all the fits performed with this method is not to estimate the actual systematic uncertainty of the final measurement, which will likely be performed on different samples and with different selection. The aim is in fact to develop methods to derive proper corrections for the simulation and to estimate the power of the various recoil definitions.

The fit range on $M^2_T$ is computed using the pseudo-data sample, i.e. with no reweighting, using the quantile techniques which has been described before. Then, for each reweighted mass, a template is filled. The normalization of the template is fixed to be the same of the pseudo-data: in fact, the maximum likelihood estimator of the histogram normalization coincides with the integral of the histogram of the data if the normalization is uncorrelated from the other fit parameters. The integral of the pseudo-data histogram is usually rescaled to the rough expected number of entries in the full sample collected by CMS at 8 TeV, in order to have a rough estimate of the statistical uncertainty obtained with the various definitions of $M^2_T$. This number is about 25 M for the $W+$ case. The $n_i^{T,D}$ are taken as the bin contents of the histograms after the various reweighting procedure, for the different masses or for injecting discrepancies between templates and pseudo-data. The uncertainty on these variables do not take into account the various weights that are applied, meaning that they are not the Poisson uncertainties on the true number of entries rescaled by the weights. Indeed it is easy to properly treat the uncertainties and to combine them in the $\chi^2$ as long as the templates are much larger than the pseudo-data. This condition would imply that, for a given bin, the uncertainty on the templates is negligible with respect to that on the pseudo-data. Since in the case under study this condition is not satisfied, and moreover, the uncertainty on templates and pseudo-data are extremely correlated, taking into account proper uncertainties on the bin contents is more difficult. Therefore, the $\chi^2$ is computed as if the templates were perfect and no reweighting was applied: this assesses correctly the expected statistical uncertainty in case of the chosen scaling of the integral of the distribution, such as the expected number of $W$ bosons present in the sample collected by CMS.
at 8 TeV. There is a residual statistical fluctuations though, that is due to the finite statistic in the sample. Indeed statistical fluctuations in the bin contents are not negligible and, when a reweighting is applied, they grow or decrease depending on the weight. This makes the bias and the statistical uncertainty fluctuate by a small amount with respect to the expected statistical uncertainty if the templates were uncorrelated from the pseudo-data, which can be assessed in the following way. The sample is split in a big number of sub-samples and the fit procedure is repeated for each of them. Then, the mean of the bias distribution is taken as the result, and the uncertainty on the mean is taken as the uncertainty on the bias estimation. This is a sort of bootstrapping method. The same procedure is used for the statistical uncertainty of the fit. A subtle point is the following: the mean of the bias distribution does not coincide with the bias obtained by performing the fit on the full sample. In fact, as the errors are not handled correctly in the $\chi^2$, the results of the fit on the sub-samples do not combine in the expected way. For the same reasons, the $\chi^2$ value at the minimum of the scan is not distributed as a $\chi^2_{\nu}$ with a number of degrees of freedom $\nu = n_{\text{bins}} - 1$. Thus no Pearson test can be made with this value.

In case the number of sub-samples in which the sample is split is small, the RMS is biased towards smaller values and fluctuate significantly. Thus, in order to solve the first problem, an unbiased estimator of the standard deviation is used:

$$\hat{s} = \left(\frac{n-3}{2}\right)! \cdot \sqrt{\frac{2}{2\left(\frac{n-2}{2}\right)!}} \left(\sum_{i=1,\ldots,n} (x_i - \bar{x})^2\right)^{1/2}$$

To solve the second problem the only way is to consider a large number of splits. In fact, assuming that the width of the distribution of the bias scales as $\sigma_n = \sigma_0 \sqrt{n}$, where $n$ is the number of splits, $\sigma_0$ is the standard deviation of the distribution of the bias for the total number of entries, that is estimated as the uncertainty on the average of this distribution. It is possible to demonstrate that $\Delta \sigma$, that is the uncertainty on the estimate of $\sigma$ is independent of $n$. Thus, the uncertainty on the estimate of $\sigma_0$ is $\Delta \sigma / \sqrt{n}$, which decreases by enlarging the number of splits. In the rest of this chapter, unless specified otherwise, $n$ is chosen to be 100.

6.3 Systematic uncertainty related to the $W$-$p_T$ spectrum

The first systematic uncertainty that has been evaluated is the one due to an imperfect knowledge of $f(z)$, and specifically what is the effect of an imperfect knowledge of the $W$-$p_T$ spectrum. There are two main reasons for doing so. First of all, as already mentioned, this is one of the biggest problems of the analysis at the moment. The most advanced computations, such as NNLO + NLLO predictions, are accurate at the ten percent level. The left plot of figure 6.1, taken from [51], shows a comparison of the $Z$-$p_T$ spectrum between the CMS 7 TeV data [52] and the most accurate predictions: in red the NLL+NNLO prediction is shown, whereas in blue the NNLL+NNLO computation is plotted. The ratio plot shows the ratio between the data and the blue histogram, together with the theoretical uncertainty (blue band). Discrepancies are of the order of 10%. Similar uncertainties are present in $W$ case, however there is no precise
experimental result in the very low $p_T$ region, capable to resolve the peak for instance, in this case. During a $W$ mass analysis, exploiting specific tunings of the MC generators based on the $Z$ measured data, it is possible to reduce this uncertainty to a few percent. Anyhow this uncertainty can result in a large systematic uncertainty when fitting $M_W$.

In this study there is no purpose of assessing the value of the systematic uncertainty due to the imperfect knowledge of the $W$-$p_T$ spectrum. In fact the CMS measurement will be probably performed with different MC data-sets and with different cuts and selections. In this case, the uncertainty on the $W$-$p_T$ spectrum will be parametrized with a set of nuisance parameters, which will be then fitted, together with the $M_W$ mass. This study instead aims at making a ranking of the various recoil definitions in terms of mitigation of the $W$-$p_T$ systematic uncertainty versus the expected statistical uncertainty. The final results of the uncertainties will be different, but the relative power will be similar to what is obtained in this study.

In order to do so, for each recoil definition, the $W$-$p_T$ spectrum in the templates is reweighted according to a certain function, in order to simulate a different $W$-$p_T$ spectrum. The applied weight is $w = 1 + \alpha g(p_T)$, where $\alpha$ is the amount of the $W$-$p_T$ variation, which is varied around few percents, in order to reproduce variations of the same entity of the uncertainties on the most accurate predictions, and $g(p_T)$ is an arbitrary function of $W$-$p_T$. What is called percent in the previous sentence is the value of $\alpha$. To compare it to the theory uncertainty, it is useful to look at the right plot of figure 6.1. It shows the ratio between the reweighted $W$-$p_T$ distribution and the nominal one, for four different values of $\alpha$ and assuming a linear function for $g(p_T)$. The effect of $\alpha = 1\%$ is a discrepancy of about $1\%$ at $p_T \simeq 0$ GeV, a null discrepancy at $p_T \simeq 12$ GeV, and of about $2\%$ at $p_T \simeq 50$ GeV. Variations of $\alpha \simeq$ few percents are of the same order of the expected uncertainty in the prediction of the $W$-$p_T$ spectrum. A linear function is used in this case because it is the simplest reweighting that can be thought of. All the following results will be obtained with this reweighting. Different reweighting procedures have been studied and the results, although different, show the same relative power between the various recoil definitions.

The fit procedure is repeated, once fixed $g(p_T)$, for all the recoil definitions under study, and for various values of the parameter $\alpha$. As an example, in the left plot of figure 6.2 the various $\chi^2$ scans are shown for the STK recoil (the TK recoil scaled by a factor 1.8), and for different values of $\alpha$: 0%, 1%, 3%, and 5%. This example shows that the position of the minimum, i.e. the bias of the fit, evolves significantly with the various values of $\alpha$, whereas this is not the case for the width of the parabola, which is related to the statistical uncertainty. This is a general properties of all these $\chi^2$ scans: thus the statistical uncertainty is taken as independent of the value of the variation $\alpha$.

For each variable, the bias as a function of the $p_T$ variation is studied. The right plot of the figure 6.2 shows the fitted bias, with the statistical uncertainty, as a function of the $p_T$ variation, for the various $M_T^2$ definitions. A fit with a straight line, with the intercept fixed at zero, is superimposed to each set of points. The linear fit is a very good description, meaning that it is possible to approximate the bias with a linear function of the $p_T$ variation. Thus the slopes correspond to the dependence of the various recoil definition on the $p_T$ spectrum: the smaller the
6.3 Systematic uncertainty related to the $W_{pt}$ spectrum

Figure 6.1: Left: comparison of the $Z_{pt}$ spectrum between the one measured by CMS [52], and the NLL+NLO prediction (red) and the NNLL+NNLO prediction (blue). The plot is taken from [51]. Right: Ratio plot between the reweighted $W_{pt}$ spectrum and the nominal one, for various entities of the reweighting. The reweighting is performed with a linear function of the $W_{pt}$.

more independent, meaning the better systematic mitigation power due to an imperfect $W_{pt}$ knowledge.

![Graph showing $W_{pt}$ spectrum variation](image)

Figure 6.2: Left: example of $\chi^2$ scans for different values of the $W_{pt}$ spectrum variation $\alpha$. These examples are obtained with the STK recoil definition, using a linear reweighting. Right: Biases, together with the statistical uncertainties, as a function of the $W_{pt}$ spectrum variation $\alpha$. The various colours refer to different recoil definitions. The superimposed straight lines are the linear fit to the points.

In order to summarize these results, the plot in figure 6.3 is helpful. It shows, for each recoil definition, or equivalently for each $M^2_T$ definition, the $W_{pt}$ mitigation power, expressed in MeV/percent, which is obtained as the slope previously fitted with a straight line, versus the expected statistical uncertainty, obtained as the width of the parabola.

The first point that should be noticed is the gen (black), close to the axis origin. This is produced
with the generator level recoil: it is not possible to find a better variable than this one. Then, there are several definitions with a statistical uncertainty between 4 and 6 MeV. Starting from the top, using the lepton $p_T$ only (red point) provides a lower mitigation power with respect to the other recoil definitions, as known. The TK recoil (blue) is what has been used in the CMS W-like analysis [31]. Scaling by the correct factor (green) reduces the systematic uncertainty by one third, and a similar result is obtained using the PK recoil (deeper brown). The MNPK recoil (lighter brown) shows a reduction of the systematic uncertainty by two thirds with respect to the TK recoil. In addition the PF recoil provides a systematic uncertainty which is comparable with the one of the MNPK recoil, but, due to its poor resolution, has a statistical uncertainty that is 60% larger than the MNPK case.

To summarize, this study results in a ranking of the recoil definitions in terms of systematic uncertainty due to the imperfect knowledge of the $W-p_T$ spectrum versus the expected statistical uncertainty. The best result obtained with the regression technique is MNPK: it provides a systematic uncertainty on the $W-p_T$ spectrum that is about one third of the one provided by the TK recoil, and a statistical uncertainty 60% smaller than the one obtained with the PF recoil. It is remarkable that this study is not assessing the statistical or systematic uncertainty on the
6.4 Systematic uncertainty related to the modeling of the recoil

The second systematic uncertainty which has been evaluated is due to the recoil mismodeling. It is induced by discrepancies between data and MC in the distribution \( f(y|z) \). Once the recoil definition is fixed, such discrepancies arise from discrepancies in the distribution of \( f(x|z) \). We recall here that \( x \) refer to the features of the recoil, \( y \) represents the variables which describe the kinematics of the recoil, such as magnitude and azimuthal angle, and \( z \) stands for the variables which describe the kinematics of the boson. The pdf \( f(x|z) \) represents a sort of response, i.e. how the recoil appears in the detector once fixed the \( W \) boson kinematics. In fact it incorporates both the hadronization of the recoil and the acceptance and resolution of the detector. The definitions that will be treated in this section are: track recoil (TK), particle flow recoil (PF) and the recoil obtained with the \( e_{1,2} \) regression (MNPK).

Discrepancies in the \( f(x|z) \) arise from an imperfect modeling of the hadronization process, as well as an imperfect description of the detector response, particularly for low momentum particles. It is expected that the more information is exploited by a given definition, the larger the systematic due to the mismodeling of the recoil will be.

There are several steps for evaluating this systematic uncertainty. First of all, discrepancies in \( f(x|z) \) and in \( f(y|z) \) distributions are analyzed for various definitions. Then the bias due to these discrepancies is evaluated with a fit procedure, where templates differ from pseudo-data only for the observed discrepancy in \( f(y|z) \). The procedure described so far is performed by comparing \( Z \) simulated and experimentally collected events. If the systematic uncertainty on the \( W \) mass is larger than the target precision, the comparison between \( Z \) MC and data is exploited in order to correct the simulation and agree with the data. This correction is subsequently brought to \( W \) simulated events. Since \( W \) and \( Z \) events are in general different, proper assumptions and corrections are made before applying this correction, the validity of which is tested at the same time. Eventually, the residual bias due to a mismodeling of the recoil is assessed, giving the size of the systematic uncertainty.

The \( z \) variables are chosen to be the boson \( p_T \) and \( p_z \), although only the first is important. In fact, the recoil depends on the \( W \) kinematics, but the recoil experimental variables show only little correlation with \( p_z \). An important assumption is that the \( y \) distribution does not change with the mass, once \( p_T \) and \( p_z \) are fixed. This is carefully tested before transporting any correction from the \( Z \) to the \( W \). It has been chosen to characterize the \( W \) longitudinal kinematics with \( p_z \) instead of the rapidity \( Y \) because of two reasons. First of all, the rapidity mixes \( p_z \) with the boson mass \( M \). Using the rapidity \( Y \) one would have introduced an artificial dependence on the mass, although probably very small. In addition, the correlation between the boson \( p_z \) and the recoil arises from the fact that, at the generator level, the recoil \( p_z \) is equal to the boson \( p_z \), but the same identity does not hold with the rapidities. Furthermore, the rapidity distributions have
Chapter 6. Systematic uncertainties

steep edges in their ranges, which depend on the boson mass. Thus, the two edges are different for Z and W, and there is a rapidity region in which there are W events but there are no Z events. Thus, calibrating the MC with Z events could be impossible in that rapidity region.

An important point is the following. From an experimental point of view, in order to evaluate the discrepancies in the conditional distribution \( f(y|z) \), it is important to make bins of the \( y \) variables. However, these bins cannot be too small because of the finite statistics of the data-sets. In the usual case of bins with non-zero width, if the underlying distributions of the \( z \) variables are different, discrepancies in the distributions of \( f(y) \) may arise even if there is no discrepancy in the \( f(y|z) \). For this reason, in the following, when comparing a certain variable between two different samples, a proper reweighting of the Z variables would be needed. All the estimates of the biases are performed through a fit, using the same sample to produce both templates and pseudo-data, and injecting a discrepancy between templates and pseudo-data only in the \( f(y|z) \), keeping the same \( f(z) \). This removes the effect of a discrepancy in the \( f(z) \) from the estimate of the bias.

It is notable that, in case of comparing the Z and W samples, the difference in the spectrum of the \( z \) variables is very large, as shown in the preliminary studies performed in section 4.1. In case of comparing the Z data and simulation, the differences, although much smaller, are not negligible, as shown in figure 6.4. This figure shows the integrated distribution of Z-\( p_T \) (left plot) and Z-\( p_z \) (right plot) as obtained in data and simulation. The Z momentum is reconstructed as the momentum of the dimuon pair. No scale correction on the muon transverse momentum has been applied. The difference, particularly relevant in the Z-\( p_T \) distribution, are of the order of 10% at \( p_T \simeq 0 \) GeV and about few percents on the whole spectrum.

![Figure 6.4](image.png)

Figure 6.4: Comparison of the Z-\( p_T \) (left plot) and Z-\( p_z \) (right plot) distributions between data and simulation. The Z momentum is reconstructed as the momentum of the dimuon pair. No scale correction on the muon transverse momentum has been applied.
6.4 Systematic uncertainty related to the modeling of the recoil

6.4.1 Morphing procedure

In order to correct the discrepancy between data and simulation and to evaluate the effect of a possible discrepancy on the $M_{W}$ measurement, a morphing procedure has been developed. For both these purposes, a way to correct the distribution $f(x|z)$ or $f(y|z)$ is needed. This can be done by reweighting the distribution of $x$ or $y$ in bins of the $z$ variables. However, a multidimensional reweighting requires large bins in order to have enough events in each of them. Thus, an unbinned morphing procedure has been developed. A morphing is a procedure that changes the value of a variable on an event by event basis. If the correction was applied on $x$, the distribution of the $y$ would be properly corrected by construction. As the $y$ space is smaller than the $x$ space, the correction can be effectively applied on the $y$ space, in a simpler way. As already mentioned, the $z$ variables are chosen to be the boson $p_T$ and $p_z$. The assumption that there is no need of adding the boson invariant mass to the list of the $z$ variables is carefully verified in section 6.4.3. The $y$ variables are instead those describing the recoil kinematics, and they have been chosen to be $p_T^h$ and $\Delta\phi_{h,W}$.

The development of such a morphing procedure, capable of morphing a 2D distribution conditionally to two more variables, is an important original part of the work which contained an important technical and mathematical part, which is explained in the following paragraphs. However, the results in terms of systematic uncertainties are not directly dependent on the morphing procedure itself: thus, the freedom of reading this part is left to the reader, who can skip it and continue with section 6.4.2.

This morphing is performed by extending the quantile morphing in N dimensions (in this case two), conditionally to k variables (in this case two), using a Boosted Decision Tree algorithm. The developed technique is an extension of the quantile morphing procedure, based on a BDT quantile regression. All these concepts are recapped in the next three paragraphs.

Quantile morphing  A morphing is a procedure to change the value of the single events of a data-set distributed accordingly to the pdf $f(x)$, in order to make it distributed accordingly to a second pdf $g(y)$. In an abstract way it is represented by a function $y = \alpha(x)$, and it is basically a change of variables. These $x$ and $y$ variables are not related to those introduced previously in section 6.1 with a physical meaning: in this paragraph $x$ and $y$ do not refer to any physical quantity, they are just used to explain the procedure.

In 1D there is a unique differentiable way to move from a distribution to another, and it is identified by the formula of the change of variables:

$$g(y) = f(x) \left| \frac{1}{\alpha'(x)} \right|$$

Unfortunately often there is no explicit expression for $f(x)$ and $g(y)$, which are taken empirically from a data-set, and thus it is not possible to identify $\alpha(x)$ in this way. Another way to get to the same result is the quantile morphing: the function $y = \alpha(x)$ is identified as the one which
preserves the cumulative distributions $F(x)$ and $G(y)$. In formulas $F(x) = G(y(x))$ for each value of $x$. It can be demonstrated that the two coincide:

$$F(x) = \int_x^\infty f(u)du = G(y) = \int_y^\infty g(t)dt = G(y(x)) = \int_{\alpha(x)}^\infty g(t)dt$$

Taking the derivatives of both terms with respect to $x$:

$$f(x) = g(\alpha(x))|\alpha'(x)|$$

This gives a procedure to morph two 1D distributions independently of the explicit expression of the pdfs. The function $\alpha(x)$ can be obtained in terms of the cumulative distributions as $\alpha(x) = G^{-1}(F(x))$, as sketched in figure 6.5.

![Figure 6.5: Sketch of the quantile morphing procedure.](image)

In the 2D case this transformation is not unique. A way to find one of these transformations is to extend the previous reasoning in 2D in the following way. First of all the 1D quantile morphing is performed on the first variable of the two. Secondly, another 1D quantile morphing is performed, conditional to the first variable. This reasoning can be easily extended to the N dimensional quantile morphing. This transformation is not unique: for instance it depends on the order of the variables. However, the transformation itself is not important: the goal is to transform the first pdf into the second pdf.

**Quantile regression** Recalling the concept explained in section 5.2.1, a regression is a technique which, starting from a certain dataset $(x, y)$, "trains" a function which is able to predict a certain parameter of the distribution $p(y|x)$. The training is a fit, which determines the parameters $a$ of a certain functional form $y = f(x; a)$, by minimizing the so called loss function, which can be taken as minus the logarithm of the likelihood. Depending on the loss function, the target of $f(x; a)$, which is the value of the pdf $p(y|x)$ which we want to regress, can be the mean, the median, or whatever parameter related to the pdf $p(y|x)$. The quantile regression minimizes a loss function so that $f(x; a)$ reproduces a certain fixed quantile of $p(y|x)$. To understand the
proper loss function we will proceed in steps. It is important to remark that these \( x \) and \( y \) variables are not related to those introduced previously in section 6.1 with physical meaning. We use in this paragraph the standard notation with \( x \) for the independent variable and \( y \) for the dependent variable in the context of the regression algorithms.

The mean square error (\( MSE \)) is the simplest loss function, and follows directly from the concept of the least squares fit. The target predicted with the \( MSE \) is the mean value of \( p(y|x) \). In fact the derivative of \( MSE \)

\[
MSE = \sum_i (y_i - f(x_i))^2
\]

is \( \sum_i (y_i - f(x_i)) \). It is equal to zero when \( f(x) \) is the mean of \( p(y|x) \).

Another well known loss function is the mean absolute error (\( MAE \)):

\[
MAE = \sum_i |y_i - f(x_i)|
\]

it is the proper loss function for regressing the median of the distribution, which coincides with the quantile \( \tau = 50\% \). This loss function is basically the sum of the absolute values of the residuals: its derivative is \( \sum_i \text{sgn}(y_i - f(x_i)) \), which is the sum of the signs of the residuals. This is equal to zero only when the number of positive residuals, that is the number of \( y \) values larger than \( f(x) \), is equal to the number of negative residuals. This is the definition of median or quantile \( \tau = 50\% \). By extending the \( MAE \) by giving a different weight to the positive and negative residuals it is possible to regress whatever quantile \( \tau \) of the distribution \( p(y|x) \). The proper weights are \( \tau \) and \( 1 - \tau \), and the resulting loss function is the following:

\[
Q_{\text{loss}}(\tau) = \sum_{y_i < f(x_i)} \tau(y_i - f(x_i)) + \sum_{y_i > f(x_i)} (1 - \tau)(f(x_i) - y_i)
\]

it is important to notice that, in case of \( \tau = 0.5 \) it coincides with the \( MAE \) apart for a factor \( 0.5 \), which is irrelevant. Its derivative is \( \sum_{y_i < f(x_i)} \tau - \sum_{y_i > f(x_i)} (1 - \tau) \). This is equal to zero only when the \( \sum_{y_i < f(x_i)} \) is the \( 1 - \tau \) fraction of the sample and \( \sum_{y_i > f(x_i)} \) is consequently the \( \tau \) fraction of the same sample, i.e. \( f(x) \) is the \( \tau \) quantile of \( p(y|x) \).

A subtlety is represented by the fact that this loss function is not convex, making the minimization process more difficult. However, while summing up on many events, the function becomes approximately convex and a minimization is thus possible.

To summarize, a regression which minimizes the quantile loss function for a given value of \( \tau \) is a quantile regression, which predicts the \( \tau \)-quantile of the distribution of \( p(y|x) \).

**Extension to \( N \) dim conditional to \( k \) dim with BDT** The goal is to perform a morphing of the 2D distribution of the recoil variables \( y \), conditionally to two variables which identify the boson kinematics \( z \). The extension of the quantile morphing to 2D is obtained by morphing the first variable, and then the second variable conditionally to the first one. Thus a 1D quantile morphing conditional to \( k \) other variables is needed.
The following algorithm has been developed in order to address the previous problem. Two sets of quantile regressions are trained, for equi-spaced values of the quantiles, such as \{0, 0.1, 0.2, ..., 0.9, 1\}. If we want to morph sample A into sample B, the first set of quantile regressions is trained on sample A and second on sample B. Each quantile regression has, as input variables, the \(k\) variables to which the morphing is required to be conditional to. Specifically, suppose to morph \(p_T^b\) as a first variable and \(\Delta \phi_{h,W}\) as the second variable. In the first morphing, the \(k\) variables will be \(\{p_T^W, p_T^L\}\), and in the second they will be \(\{p_T^W, p_T^L, p_T^b\}\). Then, for each event, meaning for each set of the \(k\) input variables, the two cumulative distributions for the sample A, called \(F(x)\), and for the sample B, called \(G(x)\), are known in an approximate way. To perform the morphing we need to compute \(G^{-1}(F(x))\): this is done by matching the points corresponding to the same quantiles of \(F\) and \(G\), and by interpolating the function between the points.

The architecture used to implement the quantile regressions is the Boosted Decision Tree (BDT), implemented through the Scikit-Learn library [53]. BDTs are, similarly to neural networks, a way to represent multidimensional non-linear functions. A more complete discussion on BDT is given in Ref. [54], [55], and [56].

**Subtleties in the evaluation of the biases with the fit** There are several sources of statistical uncertainty which can bring an additional spread on the resulting bias. The fit procedure evaluates, as shown previously, the statistical uncertainty due to finite statistic of the fitted sample. However, the morphing procedure brings two more sources of statistical uncertainty. First, the finite statistic of the sample where the morphing is trained, which thus can learn statistical fluctuations. Secondly, the finite statistics of the sample where the morphing is applied to, that thus can morph statistical fluctuations.

**6.4.2 Evaluating the effect of discrepancies on the \(M_W\) fit**

The first step to study discrepancies between the data and the simulation in the distribution of the \(f(x|z)\), and consequently \(f(y|z)\), is to look at the distributions of the \(x\) or \(y\) in bins of the \(z\) variables, for simulated and experimentally collected \(Z\) events. In figure 6.6, an example of the discrepancy in the \(x\) distribution is shown, for \(Z\) simulated events and in different bins of \(Z-p_T\) and \(Z-p_z\).

The bias due to discrepancies between data and simulation such as those shown in figure 6.6 is evaluated using the morphing procedure. A morphing is trained between the \(Z\) data and \(Z\) MC sample, for both the TK and regression recoil definitions, conditionally to the reconstructed \(Z-p_T\) and \(Z-p_z\), obtained by summing up the reconstructed lepton momenta. These two variables are in fact present in both samples and they only slightly differ from the corresponding quantities at generator level. This morphing is then applied to the MC, producing the data-like (DL) recoil. In figure 6.7 the distributions of the recoil \(p_T\), \(\Delta \phi_{h,W}\), and the corresponding \(M_T^2\) are shown for TK (first row) and MNPK (second row). In each plot the distributions in the data, in the MC before applying the morphing, and in the MC after the morphing are plotted. The agreement increases
6.4 Systematic uncertainty related to the modeling of the recoil

Figure 6.6: Data simulation discrepancy on two of the input variables of the regression: magnitude of the TK recoil (left) and number of tracks (right). The first variable is also a $x$ and $y$ variable of the TK recoil definition. These plots show the ratio data/simulation of these variables, in bins of $p_T$ and rapidity of the $Z$ reconstructed boson. The discrepancies are of the order of 10-20%.

After the morphing, and residual differences are mainly due to discrepancies in the underlying $Z$-$p_T$ spectrum. In fact these plots show the distribution $f(y) = \int dz f(y|z)f(z)$: a discrepancy between the $f(y)$ may arise from a discrepancy in the $f(z)$, even if there is no difference in the $f(y|z)$. The small discrepancy seen in the plots after the morphing may arise from the discrepancies seen in figure 6.4: a more correct comparison would require a proper reweighting of the $Z$-$p_T$ and $Z$-$p_z$ distributions. No correction for the lepton transverse momentum scale has been applied before computing the $M^2_T$.

By performing a fit using the DL variables to produce templates and the standard (non-morphed) ones to produce pseudo-data, it is possible to evaluate the effect of these discrepancies before the correction.

The result of the fit gives the bias in case where no correction is applied. The results are: $29.18 \pm 10.85$ MeV for the TK recoil, $-80.24 \pm 23.73$ MeV for the PF recoil, and $140.46 \pm 14.38$ MeV for the MNPK recoil. The three variables show different systematic uncertainties, as expected. The PF recoil is exploiting more information than the TK recoil and using neutral clusters, which are usually more difficult to calibrate. The MNPK recoil is exploiting even more information, and thus results with the largest bias before the correction, among the three variables. All the biases do not agree with zero, and thus require a proper calibration and correction.

Some more control checks are subsequently performed. In order to test if the statistical uncertainties assessed with the fit technique are reliable, and to evaluate any additional bias produced by the morphing, a self-like (SL) fit is performed. To this effect, the sample is split in three
equally populated sub-samples, and two morphings are trained between the first and the other two. The morphing is applied to the first sub sample, obtaining SL variables for the first and the second morphing. Three fits are then performed between all the three pairs of variables. The result is summed up in figure 6.8, where the distributions of the bias obtained with one hundred sub-samples are shown for the three different fits.

The average biases are: $-12.58 \pm 13.45$ MeV, $-2.65 \pm 16.84$ MeV, and $-20.19 \pm 17.60$ MeV. They agree with zero within one or two $\sigma$, confirming that there is no additional bias introduced by the morphing. The fact that the three biases are all negative is suspicious: although this is not yet a hint of bias, further investigation is probably needed. The statistical uncertainties on these biases are of the order of 10-20 MeV. These numbers are consistent with the statistical uncertainties on the biases assessed in the DL checks. The fact that these uncertainties fluctuate in an important way is a hint of the fact that the statistical uncertainties which have not been evaluated, due to the finiteness of the sample on which the morphing is trained and applied, are important and should be taken into account. Therefore, the statistical uncertainties on the biases determined

Figure 6.7: One-dimensional projection of the recoil $p_T$ (left) and of $\Delta \phi_{h,Z}$ (right), and transverse mass distribution computed with the reconstructed muon (right), in $Z+$ events. The first row is obtained with the TK recoil, whereas the second row contains the recoil definition MNPK, as obtained from the regression.

The distributions show the data (black), the MC before applying the morphing (red), and the MC after the morphing (blue). As seen in the ratio plots, after the morphing the agreement improves significantly, at a percent level.
6.4 Systematic uncertainty related to the modeling of the recoil

There are two main issues which arise when trying to apply the morphing to the W sample. First, it is not obvious that the distributions of \( f(y|z) \) are the same in Z and W samples. If the observed features of the recoil depended only on its momentum, this statement would be obvious. However, there are several subtleties which distinguish Z and W production, such as the much larger presence of heavy flavours in the Z production mechanism than in the case of the W. This assumption is in fact needed in order to apply any morphing, which has been trained on the Z samples, on the W sample. This assumption is carefully verified in the following. Secondly, in the W simulation there is no reconstructed boson, but only the generator level quantities are present. This can bring an additional bias when using the generator level quantities instead of the experimentally reconstructed ones to apply the morphing. In fact, the boson kinematics appears both in the \( z \), and in the \( y \) in order to compute \( \Delta \phi_{h,W} \) which depends on the definition of the boson.

To face the first issue, a morphing between the W and Z samples, using the generator level...
quantities for the boson, has been trained, in both directions. This morphing is complex, as there are regions of the phase space, particularly in the tails of the $p_T$ and $p_z$ distributions, where there are many events of one sample and much fewer of the other. After applying it, $Z$-like (ZL) and $W$-like (WL) variables for the recoil are produced in the $W$ and $Z$ samples, respectively. In figure 6.9 the comparisons are shown for the MNPK recoil for the ZL (first row) and WL (second row) case. The plots, similar to those seen in figure 6.7, shows the distributions of $p_T$ and $\Delta \phi_{h,W}$ for the $W$, $Z$, and ZL or WL, with the corresponding ratio plots. The cut on the boson kinematics, $7 \text{ GeV} < p_T < 11 \text{ GeV}$ and $|p_Z| < 100 \text{ GeV}$ (which corresponds to a rapidity of $Y = 0.95$ in the $Z$ case, and $Y = 1.03$ in the $W$ case) has been applied in order to reduce discrepancies due to the difference in the $z$ variables between $W$ and $Z$. This effect is similar to the one seen in figure 6.7. However, the difference between $Z$ and $W$ are much larger than those between $Z$ data and $Z$ simulation. A cut in the region where the distributions are similar is thus needed. However, a more fair comparison would require a reweighting of the distribution of the $z$ variables in order to cancel this effect. In both the ZL and WL cases the distributions agree fairly well and no major difference can be seen between the original variables and the ZL or WL. This is a hint for the fact that there is no difference between $Z$ and $W$, and thus the morphing is only transforming statistical fluctuations.

The effect of this possible discrepancy is then evaluated with a fit procedure, using the WL (ZL) variables to produce the templates and the variables before the morphing for pseudo-data. This check has been performed only with the MNPK recoil definition. The result is $1.69 \pm 11.58 \text{ MeV}$ for the WL check, and $-36.24 \pm 13.33 \text{ MeV}$ for the ZL check. The ZL cross check does not agree with zero bias, whereas the WL does. This might be due to the differences in the distribution of the $z$ variables. Such differences could in fact make the morphing difficult to train and apply, as previously alluded to. For example, in case one of the two samples contains much fewer events at a certain value of $z$ than the other, training and applying the morphing in that region will be difficult. This is due to the fact that the quantile regressions, for the sample with fewer events, will converge worse than for the sample with more events. Further investigation is thus needed.

The second problem, which is related to the fact that no reconstructed boson is present in the $W$ MC sample, has been overcome in the following way. The morphing trained between the $Z$ samples, which implements the data-MC corrections, is applied on the same sample using generator level quantities instead of reconstructed quantities. This produces $\text{gen-reco}$ variables, that are used to generate templates, in order to fit the pseudo-data produced with the standard variables. This check, performed both with TK and MNPK recoil, results in the following biases: $-5.34 \pm 7.19 \text{ MeV}$ for the TK recoil and $-12.01 \pm 11.34$ for the MNPK recoil. The result seems to be almost compatible with zero within the uncertainty. Further investigation is probably needed in this case too.

In case this bias were important, a possible solution would be to compute a sort of reconstructed boson in the $W$ simulated events. This can be done by applying a smearing of $\approx 2\%$ on the generator level neutrino transverse momentum and summing up this momentum with the reconstructed lepton. It is not important that this technique is perfect: it is enough that the resulting
6.4 Systematic uncertainty related to the modeling of the recoil

Figure 6.9: Distributions of the recoil $p_T$ (left) and $\Delta\phi_{h,W}$ (right), for the MNPK recoil. The first row, show the distributions in the $Z$ sample (black), $W$ sample (red), and $W$ ZL (blue), whereas the second row shows the opposite. A cut on the $z$ variables ($7 \text{ GeV} < p_T < 11 \text{ GeV}$ and $|p_Z| < 100 \text{ GeV}$) has been imposed in order to mitigate the effect of the difference in the $z$ distribution between $Z$ and $W$.

systematic uncertainty is small.

In all of these cases, if the resulting bias is compatible with zero within the statistical uncertainty, the systematic uncertainty that must be quoted is this statistical uncertainty. It can be seen that the main limitation of this analysis so far is the statistical power of the sample, which results in quite large statistical uncertainties, of the order of tens of MeV.
6.4.4 Residual data-MC discrepancy after the correction

In order to evaluate the systematic uncertainty due to possible residual discrepancy between the data and simulation after the morphing, the following method is used. On the $Z$ sample, a new morphing between the original sample and the morphed sample is trained. When applying it, the new variables $\text{data-like 2} (\text{DL2})$ are obtained. The fit performed using DL2 to generate templates and DL for pseudo-data assesses the residual bias. The results are $-14.39 \pm 6.57$ MeV for the TK recoil and $-11.22 \pm 9.83$ MeV for the MNPK recoil. Neither of the two agrees with zero within this estimate of the statistical uncertainty, although the disagreement is small. Further investigation with larger samples is needed in this direction too, in order to understand if the biases agree with zero within a smaller and more precise estimate of the statistical uncertainty, or if a residual bias, which can be estimated and corrected, is still present.

Once again, in case this checks showed a bias in agreement with zero within its statistical uncertainty, the statistical uncertainty itself would assess the residual bias.

6.5 Conclusions

In the course of this chapter several results have been achieved.
First, the origin of the various systematic uncertainties related to the recoil has been understood, in terms of the factorization approach.
Secondly, a simplified version of the template fit to the $M_{T}^{2}$ distribution has been implemented. This simplified version is meant to estimate the expected statistical uncertainty on the $M_{W}$ extraction using the CMS 8 TeV dataset. Additionally, the fit is used to assess the systematic uncertainties due to an injected discrepancy between templates and pseudo-data.
With this technique, the bias due to an imperfect knowledge of the $W$-$p_T$ spectrum has been evaluated, providing a ranking of the different recoil definitions. The recoil definition based on the regression has shown the best result among the variables under study, reducing the systematic uncertainty due to an imperfect knowledge of the $W$-$p_T$ spectrum by a factor three with respect to the TK recoil, and the statistical uncertainty of 60% with respect to the PF recoil.
Furthermore the systematic uncertainty due to a mismodelling of the recoil has been studied. Thanks to a morphing technique which has been developed for the purpose of morphing the 2D distribution of the recoil, conditionally to the boson kinematics, the problems of correcting data/MC discrepancies and assessing this systematic uncertainty have been discussed. All the checks which have been performed are summarized in table 6.1. The final result (data-like2), which provides an assessment of the residual systematic uncertainty after the correction, is then limited by the finite size of the sample. It is in fact not perfectly compatible with zero, for both the TK and the MNPK recoil, within an uncertainty of about 10 MeV. Further studies with larger samples are needed in order to precisely assess the presence of a possible residual bias which could be eventually corrected.
Table 6.1: Summary of the control checks performed in the course of this section. Each check reports the sample on which has been performed, the recoil definition, and the variables used for the templates and for the pseudo-data. The check estimates the bias due to the injected discrepancy between templates and pseudo-data. The uncertainty on the bias is estimated by splitting the sample in 100 sub-samples and repeating the fit for each sub-sample.

<table>
<thead>
<tr>
<th>Check</th>
<th>sample</th>
<th>recoil</th>
<th>template</th>
<th>pseudo-data</th>
<th>bias [MeV]</th>
<th>uncertainty [MeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>data-like</td>
<td>Z</td>
<td>TK</td>
<td>DL</td>
<td>STD</td>
<td>29.18</td>
<td>10.85</td>
</tr>
<tr>
<td>data-like</td>
<td>Z</td>
<td>PF</td>
<td>DL</td>
<td>STD</td>
<td>-80.24</td>
<td>23.73</td>
</tr>
<tr>
<td>data-like</td>
<td>Z</td>
<td>MNPK</td>
<td>DL</td>
<td>STD</td>
<td>140.46</td>
<td>14.38</td>
</tr>
<tr>
<td>self-like</td>
<td>W</td>
<td>MNPK</td>
<td>SL1</td>
<td>STD</td>
<td>-12.58</td>
<td>13.45</td>
</tr>
<tr>
<td>self-like</td>
<td>W</td>
<td>MNPK</td>
<td>SL2</td>
<td>STD</td>
<td>-2.65</td>
<td>16.84</td>
</tr>
<tr>
<td>self-like</td>
<td>W</td>
<td>MNPK</td>
<td>SL2</td>
<td>SL1</td>
<td>-20.19</td>
<td>17.60</td>
</tr>
<tr>
<td>W-like</td>
<td>Z</td>
<td>MNPK</td>
<td>WL</td>
<td>STD</td>
<td>1.69</td>
<td>11.58</td>
</tr>
<tr>
<td>Z-like</td>
<td>W</td>
<td>MNPK</td>
<td>ZL</td>
<td>STD</td>
<td>36.24</td>
<td>13.33</td>
</tr>
<tr>
<td>gen-reco</td>
<td>Z</td>
<td>TK</td>
<td>gen-reco</td>
<td>STD</td>
<td>-5.34</td>
<td>7.19</td>
</tr>
<tr>
<td>gen-reco</td>
<td>Z</td>
<td>MNPK</td>
<td>gen-reco</td>
<td>STD</td>
<td>-12.01</td>
<td>11.34</td>
</tr>
<tr>
<td>data-like2</td>
<td>Z</td>
<td>TK</td>
<td>DL2</td>
<td>DL</td>
<td>-14.39</td>
<td>6.57</td>
</tr>
<tr>
<td>data-like2</td>
<td>Z</td>
<td>MNPK</td>
<td>DL2</td>
<td>DL</td>
<td>-11.22</td>
<td>9.83</td>
</tr>
</tbody>
</table>
Conclusions and future perspectives

These nine months of working on the recoil have resulted in some interesting results and have opened new possibilities.

First of all, the hadronic recoil represents a key ingredient of the $W$ mass measurement. In fact the main theoretical problem at this stage is the correct prediction of the $W$-$p_T$ spectrum, which has large uncertainties. However, the experimental measurement of the hadronic recoil can fight this problem: a better and better recoil definition makes the transverse mass less and less dependent on the $W$-$p_T$ spectrum.

Secondly, studies of the features of the recoil, its production mechanism, and its contents in terms of particles have been performed, leading to the guidelines for a good recoil definition. A new method, based on a semi-parametric regression, has been developed, in order to exploit as much as possible the information collected by the detector. This method has led to a new and innovative definition of the recoil.

The performance of the various definitions have been evaluated in terms of systematic uncertainties. Firstly the various sources of uncertainty related to the recoil have been listed. Thanks to the factorization approach they can be treated as uncorrelated systematic uncertainties. Furthermore, in order to study the effect of these uncertainties for the different recoil definitions, a simplified version of the template fit has been developed.

With this tool, the uncertainty related to an imperfect knowledge on the $W$-$p_T$ spectrum has been studied: a ranking of the various variables in terms of systematic uncertainty versus expected statistical uncertainty showed that the recoil definition based on the regression performed three times better than the track recoil in terms of systematic uncertainty and 60% better than the particle flow recoil in terms of statistical uncertainty.

Additionally, thanks to the development of a morphing technique, the bias related to a mis-modeling of the recoil has been studied. The effect without applying any correction on the MC results to be too large for all the recoil definitions under study. Specific corrections have been derived between the $Z$ data and MC samples, and then applied on the $W$ MC: this operation required some assumptions which have been carefully studied. The residual biases after applying the correction presents a small disagreement with zero, within uncertainties of about 10 MeV, for both the TK and the MNPK recoil.

The open points, on which future work can continue, are the following. To begin with, as the recoil definition is important for its power of mitigating the systematic uncertainty due to an
imperfect knowledge of the $W-p_T$ spectrum, a method to evaluate this power is important. The only explicit way to quantify this effect which has been found is through the fit procedure. However, understanding better which feature of the recoil makes the transverse mass almost independent on the $W-p_T$ spectrum could be important for two reasons: it could be a useful handle to improve the recoil definition, and it provides a way to evaluate the recoil definition independently on the details of the fit.

Additionally, an important limitation is the statistics of the MC sample. In fact, the estimates of the biases is limited by the statistical uncertainty with which they are assessed. So far, the residual biases due to a mismodeling of the recoil present small disagreements with zero with an uncertainty of the order of 10 MeV. Further studies performed with larger samples could assess the presence of a residual which can be eventually corrected.

Moreover, all the study performed on the biases due to a mismodeling of the recoil should be performed by applying proper reweightings, in order to cancel the discrepancy due to the underlying boson $p_T$ and $p_z$ distributions, and to better understand the discrepancy in the modeling of the recoil, i.e. in the distribution $f(y|z)$.

The problem of the possible discrepancy between $Z$ and $W$ in the modeling of the recoil is not perfectly understood. New methods for showing that there is no discrepancy or to assess it are thus needed.

Furthermore, the regression technique, which has been developed for the recoil definition, provides more information than what has been exploited so far. In fact it gives the distribution of the true gen-level $p_T$ on an event-by-event basis. This information could be in principle used for a more complicated fit to $M_W$. Moreover, it opens the possibility of a precise measurement of the $W-p_T$ spectrum in the low momentum region.
Appendix A

Semi-parametric regression as a possibility for a precise measurement of the $W-p_T$ spectrum

In the course of chapter 5, the development of a semi-parametric regression for the recoil study has been described. The semi-parametric regression predicts the distribution $p(y|x)$ on an event by event basis. In this appendix we will focus on the magnitude of $p_T$ of the recoil $p_h$, and we will forget the angle. As the recoil transverse momentum is equal to the $W-p_T$, the regression provides on an event by event basis with the distribution of the true $W-p_T$ corresponding to the reconstructed $x$ variables. This distribution has been used for extracting an estimator of the recoil on an event-by-event basis, but also to make some control checks for the convergence of the algorithm, as seen in figure 5.12, 5.13, and 5.14.

Particularly, the left plot in figure 5.13 shows a nice agreement between the true gen-level distribution and the sum of pdfs plot. This agreement opens the possibility to measure the distribution of the $W-p_T$, as with this control check we are able to reproduce it. The regression provides with a sort of unfolding of the $W-p_T$ distribution.

There are several subtleties in doing it. The major one is that the check has been made on the same MC where the regression was trained. In the case of a measurement this sum-of-pdfs procedure will be performed on the data. Thus there is an important need for a reliable MC. Particularly, in order to properly work, a MC with the correct $W-p_T$ distribution is needed. In fact, as explained in section 5.2.3, the regression depends on the $W-p_T$ spectrum present in the training sample. As we want to measure it, an idea to overcome this problem is to perform an iterative procedure, which assumes that the difference between the MC and the true is small. The regression is trained on MC, and predicts the $W-p_T$ distribution on data: this prediction is used for a new training, and the process is repeated until convergence. It is also important that the MC is reliable for what concerns the distribution of the $x$ variables given the true recoil. The morphing techniques developed in chapter 6 can work for this purpose. It is noteworthy that in
this case, there is no need for many variables which can identify well the recoil: the importance is that the functional form chosen to fit the pdf $p(y|x)$ with the semi-parametric regression is describing well the distribution in the whole phase space, and that the distribution of the input variables $p(x|p_T^{W})$ given $W-p_T$ is the same in data and MC.

No complete study on the feasibility of this measurement has been performed. However it is of interest to proceed in this way, as it can open the possibility for a new measurement. The most precise measurement of the $W-p_T$ at the LHC is performed with big bins, the first one is large 7 GeV, with no possibility for resolving the peak of the distribution. This measurement would be extremely useful in constraining the QCD NNLO+NLLO calculation, in view of a more precise measurement of the mass of the $W$ boson.
Appendix B

Neural networks settings for the recoil regression

The regression developed for the recoil measurement is based on a semi-parametric regression which has been implemented on a deep neural network architecture, as illustrated in section 5.2. The basic concepts of the neural networks have been explained in subsection 5.2.2. This appendix is meant as a list of the settings of the neural network which has been used for the $e_{1,2}$ regression, illustrated in section 5.3.

The training is performed with the Keras library, a Python wrapper for TensorFlow [50], which provides optimal support for the GPU. A complete explanation of the options and settings in the following can be found in Ref. [49].

Two similar networks for the two regressions (the first targeting the $\log(e_1)$ and the second targeting $e_2$): the first one consists of four parallel networks and the second of five parallel networks, one network for each parameter of the pdf $p(y|x)$ which is fitted. The four (five) networks share the input layers, where the input variables are provided, and have an output layer with only one neuron each, whose output is the predicted value of the parameter related to that network. The chosen architecture have three hidden layers, consisting of 100, 50, and 8 neurons. The activation function which has been chosen for the hidden layers is the *Leaky ReLU*, defined as:

$$
\phi(x) = \begin{cases} 
  x & \text{for } x > 0 \\
  \alpha x & \text{for } x < 0 
\end{cases}
$$

where $\alpha$ is a parameter of the function which has been fixed to $\alpha = 0.2$. For the output layers: for the $\mu$ parameter of the $\log(e_1)$ regression a linear function is chosen, whereas for the same parameter of the $e_2$ regression a sigmoid scaled between $-\pi$ and $+\pi$ is used. For the parameters $\sigma$, $\alpha_1$ and $\alpha_2$ of both regressions an exponential activation is used, in order to produce only positive values. For the constant parameter of the $e_2$ regression a Sigmoid scaled between 0 and $1/(2\pi)$ is used.

The weights of the neurons are initialized with the *Glorot Normal* option, whereas the biases are
initialized with the *Glorot Uniform* option.

For each semi-parametric regression a total number of about 22k of free parameters is used.

The training is performed with 100 epochs, using a batch size of 200 events. The sample has
been split into two subsamples: 70% of the events are used for the training and the rest as a
testing sample. The training sample is reshuffled at the end of each epoch. The optimizer which
has been chosen is adam [57] with a decay of learning rate of $1 \cdot 10^{-6}$ and a clip norm set to 1.
Additionally, two callback are set during the training: *EarlyStopping* with patience 8 on the
loss function of the testing sample, and *Reduce Learning rate on plateau*, with patience 3 and
reduction 0.3 on the training loss function.

Although an hyper-parameter scan has not been performed, several settings have been explored,
leading to the options that have been described.
Appendix C

Breit-Wigner reweighting

In this type of analysis, where a template fit is performed, a great number of templates is needed, each of them generated with different value of $M_W$. However, generating a whole MC sample for each value of the mass is extremely consuming in terms of computational power. For this reason, proper reweightings can be applied in order to provide the expected distribution of the variables of interest, the lepton $p_T$ or the transverse mass in the case under study, for different values of $M_W$, starting only from one generated sample.

One of these methods is based on the Breit-Wigner distribution: thanks to the finite width of the $W$ boson, which is around 2 GeV, while simulating events at a certain value of $M_W$, a lot of events with larger or smaller masses are produced as well, accordingly to the Breit-Wigner distribution

$$f(M; M_0, \Gamma) \propto \frac{1}{(M^2 - M_0^2)^2 + M_0^2 \Gamma^2}$$

where $M_0$ and $\Gamma$ are the central value and the width of the distribution, and $M$ is the mass of the generated boson, i.e. the invariant mass of its decay products.

Then, when generated events with a certain value $M_W = M_0$, and events at $M_W = M_1$ are needed, the following weight is applied:

$$w(M) = \frac{f(M; M_1, \Gamma)}{f(M; M_0, \Gamma)} = \frac{(M^2 - M_0^2)^2 + M_0^2 \Gamma^2}{(M^2 - M_1^2)^2 + M_1^2 \Gamma^2}$$

A first subtlety is that this is valid only for $|M_1 - M_0| \lesssim \Gamma_W$, i.e. the weights should be numbers close to one. Furthermore, this distribution should be convoluted with the partonic luminosities, obtained from the parton distribution functions, which modifies the Breit-Wigner shape. However, most of this effect cancel out when taking the ratio, and it is thus negligible.
Bibliography


List of Figures

1.1 The particles of the standard model: in purple the quarks are shown, in green it is possible to see the leptons, whereas the gauge bosons appear in the red boxes. 2

1.2 A sketch of the Higgs potential, usually called Mexican hat potential, due to its shape. The minimum of the potential is represented by continuum values, hence there is a continuous infinity of degenerate vacuum states, distinguished by the value of the phase. The underlying symmetry is exact though, and all the vacua are in fact physically equivalent. The choice of one of the vacua breaks the symmetry, and the Higgs field can be rewritten with a physical Higgs boson $h$, that represents the excitation along the "radial" axis, and three $\pi_i$ Goldstone bosons. The $\pi_i$ are mass-less excitations: schematically, they are moving along the continuum minima of the potential, and thus they do not require additional energy as a massive excitations. 6

1.3 The two plots show a comparison between the prediction and the experimental measured values of the top quark mass (left) and Higgs boson mass (right) as a function of the year. Even before their discovery, the EW fit was able to predict their masses with good precision and accuracy. Both plots taken from [13], where more information about the EW fit can be found. 10

1.4 The 2D plots show the probability contours resulting from the EW fit for the variables $M_W$ vs $\sin^2 \theta_{eff}^{\text{lept}}$ and $M_W$ vs $m_t$ for the left and right plots, respectively. The various contours are obtained by removing subsequently some of the measurements from the likelihood of the fit. The gray bands represent the experimental measurements. The prediction of the fit is in agreement with the experimental measurements, within $\simeq 1\sigma$. Plot taken from [1]. 12

2.1 Left: an illustration of the LHC acceleration complex. Right: the LHC rings: the four detectors are positioned in correspondence of the crossing point of the two beams. 14
2.2 Delivered luminosity by the LHC as a function of the day of the year, for the various LHC runs. As it is possible to see, during 2016 the delivered luminosity exceeded 40 fb$^{-1}$, by far larger than the expectation. This was possible thanks to the achievement and following overtaking of the instantaneous design luminosity.

2.3 An illustration of the various sub-detectors inside CMS.

2.4 An illustration of the CMS tracking system. The acronyms TIB, TID, TOB, and TEC stand for "tracker inner barrel", "tracker inner disks", "tracker outer barrel", and "tracker endcaps", respectively.

2.5 Left pane: total thickness of the CMS tracking system expressed in radiation lengths, as a function of the pseudorapidity (the material budget is almost flat in the azimuthal $\phi$ coordinate). The meaning of the acronyms is the same as for figure 2.4. This figure is taken from [18]. Right pane: tracking efficiency for charged hadrons as a function of the transverse momentum. The full offline reconstruction (red line) increases significantly the efficiency with respect to the first iteration (black line). This figure is taken from [19], where it is possible to find more details about this methodology.

2.6 Left pane: the relative transverse momentum resolution, as a function of the transverse momentum, is shown for single, isolated muons. The barrel, transition, and endcap regions, defined by $\eta$ intervals of 0–0.9, 0.9–1.4, and 1.4–2.5, respectively. The solid (open) symbols correspond to the 68% (90%) intervals. Right pane: the resolution on the measurement of the track parameter $z_0$, as a function of the pseudorapidity, is shown for single, isolated pions. The solid (open) symbols correspond to the 68% (90%) intervals. Both figures are taken from [18].

2.7 An illustration of the interactions of different particles with the various sub-detectors inside CMS.

3.1 W mass measurements performed by LEP (all the four experiments), Tevatron (CDF + DØ) and LHC (ATLAS) compared to the precision of the prediction of the EW fit (Ref. [1]). The precision reached by the ATLAS and CDF measurements is 19 MeV in both cases, and these represents the most precise measurements of the $W$ mass performed up to now. This plot is taken from Ref. [26].

3.2 The two lowest order diagrams for the $W$ pair production at $e^+e^-$ colliders. $s$-channel exchange of a photon or a $Z$ boson (left) and $t$-channel exchange of an electronic neutrino (right).
3.3 Left plot: cross section for the production of a WW pair at LEP as a function of
the center-of-mass energy. The characteristic turn-on around twice the W mass is
illustrated for three different values of the W mass. Figure taken from Ref. [28].
Right plot: invariant mass of the reconstructed W bosons from events tagged
with a four-jets final state. The data shown in this histogram was collected by the
ALEPH experiment. Figure taken from Ref. [29].

3.4 W production at hadron colliders, followed by a leptonic decay. The Feynman
diagram represents only the hard process that involves quarks, the rest of the
proton can be factorized from this hard process.

3.5 Dijet invariant mass, selected in a boosted topology with transverse momentum
between 700 and 800 GeV. The peak due to the vector bosons is clearly visible,
but it is not possible to disentangle the two contributions from W and Z bosons
because of the poor resolution. The plot shows CMS data collected during 2016,
at $\sqrt{s} = 13$ TeV, and it is taken from [32].

3.6 Left plot: Lepton transverse momentum distribution: the two cases of absent
(present) $W_p_T$ spectrum are plotted in black (blue). The effect of the $W_p_T$
is basically to smear the Jacobian peak at half of the W mass. The green line,
superimposed on the peak, indicates $M_W/2$. Right plot: Transverse mass square
distribution. In black (blue) the distribution is shown in case of absent (present)
$W_p_T$. The two distributions look pretty similar, confirming the weak dependence
of the transverse mass on the $W_p_T$ spectrum. The green line, superimposed on
the peak, indicates $M_W^2$. The red histogram shows the effect of choosing only
the charged tracks inside of the detector acceptance for computing the hadronic
recoil. Both plots show events simulated with Pythia8 standalone (for further
details about this software see Ref. [33]), with no selection applied on the events.

3.7 $W_p_T$ distribution for simulated events with Pythia8 standalone (Ref. [33]), with
no selection applied. Although the actual distribution might be slightly differ-
et, the main features are conserved. The spectrum peaks at about 4 GeV and
decreases with a long tail at high values, with a mean of about 20 GeV.

3.8 Profile plot of the $W^+$ and $W^-$ transverse momentum as a function of its rapidity
$Y$, at generator level, with no cuts due to detector effects or event selection.
The events are generated with Pythia8 standalone (Ref. [33]). The mean is
shown, together with the statistical uncertainty due to the finite MC sample that
is negligible. The behaviour is similar for the two charges, and it shows a flat
distribution for central values of the rapidity, that goes to zero while approaching
the edge of the rapidity distribution: in this case, one of the two quarks has $x \approx 1$,
and there is no energy available to produce W transverse momentum.
3.9 $p_T$ vs $\eta$ distribution for leptons from $W^-$ decay with negative helicity in the left plot, and positive helicity in the right plot. The differences in this distribution between the two helicity states may open the possibility to a precision measurement of the rapidity distribution for the two helicity states, as better discussed in Ref. [39]. The plots are taken from this reference. . . . . 40

3.10 Missing transverse momentum distribution in events selected with one and only one isolated muon with $p_T > 24$ GeV and $|\eta| < 0.2$. The missing transverse momentum $E_T$ is computed summing up all the particles reconstructed in the event. The $W$ production is the main contribution, but the background is important (around 20% of the events). A variable like $E_T$ is able to separate the contributions from the background: typically in $W$ mass analyses a lower bound on $E_T$ is required, in order to obtain a purer sample. This plot taken from [41]. . . . . . . . . 42

4.1 $p_T$ spectrum for $W$ (left) and $Z$ (right) simulated events. No event selection has been applied. The plots show the normalized distributions for the different flavours participating in the production. Contributions to the total rate that are smaller than 1% are removed from the plot, in order to make the plot more readable. . . . . 47

4.2 Left: distribution of $p_T/M$ for $W$ and $Z$ simulated events, for the two cases of light flavours only (green for the $W$ and blue for the $Z$) and all the flavours (black for the $W$ and red for the $Z$). The distributions agree qualitatively. Right: ratio plots of the $W$ and $Z$ distributions shown in the left plot. The behaviour is similar for the two cases of light flavours only (red) and all the flavours (black), and show discrepancy of the order of 10% from the unitary ratio. . . . . . . . . . 48

4.3 Rapidity spectrum for $W^+$ and $W^-$ with no cut applied (left) and with kinematic cuts on the lepton (right). The cut applied are $p_T > 30$ GeV and $|\eta| < 2.4$, and are applied on the generator level quantities. The cuts modify the shape of the distributions, particularly reducing the range of the rapidity spectrum. . . . . . 48

4.4 Left: sketch of a $W \rightarrow l\nu$ event. The particles going to the bottom represent the hadrons produced by ISR, whereas those going close to the beam axis are meant to be the proton remnants. Right: transverse view of a simulated $W$ event with the CMS detector. The yellow line corresponds to the high $p_T$ muon coming from the $W$ decay, whereas the other particles belong to the hadronic recoil. . . . . . 49

4.5 Left: number of particles belonging to the hadronic recoil for various "experimental" definitions, taking into account the acceptance of the detector in $\eta$ and in $p_T$. The average number of particles reduces significantly tightening the cuts. Right: reconstructed recoil component along the lepton direction versus the same quantity at generator level. The correlation between the two quantities is important even if the various cuts are applied, making worth to study the $p_T$ of the recoil. . . . . . . . . 50
4.6 Effect of the event selection on the transverse mass distribution. Each plot shows the distributions of the transverse mass computed in various ways: generator level quantities in blue and red, TK recoil\(^1\) in purple, no recoil in black, and the two first order approximations, in green and brown. Moving from left to right, several cuts are progressively applied. From left to right: no cut, \(|\eta| < 2.4, p_T > 30\) GeV, recoil TK < 30 GeV. The \(M_T\) distribution is sculpted: the events on the tail are affected in a more important way than the peak, making it sharper.

4.7 \(M_T^2\) distribution for \(W\) simulated events, using different definitions of the objects. In the legend, the definition of the objects is indicated: \(m\) stands for the muon and \(r\) stands for the recoil. The suffixes MC, MC_postFSR, RECO, TK stands for generator level, generator level post FSR, reconstructed level, tracks, respectively. 1\(^{st}\) is added to indicate that the first order approximation of the \(M_T^2\) has been used.

4.8 Left: distribution of the first order approximation of the transverse mass in bins of the lepton \(p_T\). The transverse mass is computed using the reconstructed lepton and the generator level value of the recoil. Right: zoom of the previous distribution in the right tail of the peak. The superimposed dashed lines identify the positions of the values where the distribution of \(M_T^2\) in the bin with 40 GeV < \(p_l\) < 41 GeV (blue histogram) takes the values corresponding to 90\% and 10\% of the peak value.

4.9 Left: \(\chi^2\) scans related to the template fits to the \(M_T^2\) distributions for the different bins of the lepton \(p_T\). Right: statistical uncertainty on the \(M_W\) extraction, based on the \(M_T^2\) template fit, as a function of the lepton \(p_T\) bin. The different curves and colours refer to the \(\sigma\) of the injected Gaussian recoil resolution.

5.1 Transverse mass distribution in CMS MC \(W^+\) events, with the reconstructed lepton, in bins of \(W-p_T\). The various colours refer to different definitions of the recoil. In blue the MC-truth recoil is plugged; in green the true angle, and the magnitude obtained by the sum of tracks is plugged; the black one contains instead the true magnitude and the reconstructed angle; whereas the red shows the transverse mass with the reconstructed recoil.

5.2 Left: 2D distribution of \(\eta\) (y axis) versus \(\phi\) (x axis) for reconstructed charged pions in \(Z^+\) events in the data. The cut \(|\eta| < 2.4\) has been imposed in order to define well the acceptance in the data and in the simulation. This makes the analysis insensitive to possible discrepancies of the tracker position between the data and the simulation. Right: Distribution of \(\Delta_{xy}\) versus \(\Delta_z\) for charged tracks in \(Z^+\) events with no simulated pile-up. \(\Delta_{xy}\) and \(\Delta_z\) are defined as the distances on the transverse plane and along the beam axis, respectively, between the PV and the point of closest approach of the track to the beamline.
5.3 The two plots show the distribution of the resolution of TK recoil in simulated $Z^+$ events, for two different reference frames: the fixed x and y axis of the detector (left) and the parallel and perpendicular axis with respect to the true recoil direction (right). The shapes of the two distributions are extremely different, showing an almost spherical distribution in the first case, and a distribution peaked at 7 GeV in the parallel direction in the second case. 

5.4 Sketch of a situation in which the angle between the reconstructed TK recoil and the true one (called MC) is large enough so that scaling it in order to have the correct magnitude (STK) is not an obvious improvement. In fact the length of the dashed red line is comparable or even larger than the dashed green line. 

5.5 Recoil resolution expressed as $|\Delta \vec{h}|$ (left) and $\Delta M_T^2$ (right), for various recoil definitions obtained by scaling the TK recoil by a constant factor, in simulated $Z^+$ events. The factor 3 clearly worsens the resolution, whereas the factors 0.5 and 1 seem to show a slightly better resolution with respect to the factors 0 and 1.6, though there is no manifest improvement. 

5.6 Left: 2D distribution of the variables x and y: y is Gaussian distributed, with mean and sigma dependent on the x variable. The dependence of the mean and sigma on the x variable is shown in the right plot. 

5.7 Left: example of a DNN with four layers. All the connections between the neurons are shown. Right: sketch of the workflow of a neuron. The inputs $x_i = 1, ..., N$ are multiplied by the weights $w_i$ and then summed up with a bias $b$. An activation function is then applied on the result of the previous operation, producing the output of the neuron. 

5.8 As the correction that must be applied to the reconstructed recoil ($\vec{h}_{TK}$) in order to get the true one ($\vec{h}_{MC}$) is a 2D vector, it can be parametrized in several ways. Two possibilities studied in the course of this thesis are shown: $c_{1,2}$ (left) and $e_{1,2}$ (right). 

5.9 Distribution of the correction coefficients for the various parametrizations present different distributions, that make them more or less suitable for a semi-parametric regression. Specifically, in this plot, the distributions for $c_1$, $1/c_1$, $e_1$, and $\ln(e_1)$ are shown. 

5.10 $e_2$ distributions for TK and PF recoil, as computed in $Z$ simulated events. The former has a narrower distribution than the latter, meaning that on average the angle is better measured. Since the angle is expected to be more complicated to be corrected, it has been chosen to start from TK.
5.11 Distribution of the log($e_1$) in bins of the magnitude of the TK recoil. The distribution of this correction varies significantly between the various bins: in the case of small recoil (black curve), it is really uncorrelated with the true one, resulting in a correction with a broad distribution. In case of large reconstructed recoil (green curve), that corresponds to the jet regime, the correction mainly takes into account the lack of the neutral particles. 

5.12 Sum-of-pdfs closure plots for the $e_1$ regression (left) and for the $e_2$ regression (right). The agreement is fairly good, within discrepancies of the order of 10%. The distribution of $e_2$ should be asymmetric as it is integrated over all the events, however it is not, because the actual target is $e_2 \cdot \text{sgn}(\phi_{NTNPV} - \phi_{TK})$ instead of $e_2$, as mentioned in 5.3.2.

5.13 Sum-of-pdfs closure plots for the magnitude (left) and angle (right) of the recoil. The agreement between the sum-of-pdfs distributions (blue) and the generator level quantity ones (black) is fairly good, with discrepancies at the level of few percents. Such discrepancies reflect what seen in figure 5.12. The distribution of STK is also shown in the plot of the magnitude of the recoil, and the distribution of TK is shown in the plot of the angle. These plots highlight the fact that the new method based on the regression is able to reproduce the $W$-$p_T$ spectrum in a better way than using only the TK recoil.

5.14 Sum-of-pdfs closure plots for the two Cartesian components of the recoil, $p_x$ (left) and $p_y$ (right). The two plots are very similar and show similar aspects to what already seen in figure 5.13.

5.15 Residual corrections on the log($e_1$) (scale factor, left) and on $e_2$ (angular correction, right), for the different recoil definitions.

5.16 Left: integrated distribution of the magnitude of the recoil for the different definitions. Right: recoil resolution expressed as $|\Delta \vec{h}|$ for the different definitions under study.

5.17 Comparison of the various definitions in terms of the residual corrections on the x (left plot) and y (right plot) axes, called $\Delta_x$ and $\Delta_y$.

5.18 $M_T^2$ resolution for the various definitions that have been studied. As the width of this distribution is expected to be related to the statistical uncertainty in the $M_T$ fit, the various definitions are expected to result in similar statistical uncertainties, a part for PF that shows a far larger width than the others.

6.1 Left: comparison of the Z-$p_T$ spectrum between the one measured by CMS [52], and the NLL+NLO prediction (red) and the NNLL+NNLO prediction (blue). The plot is taken from [51]. Right: Ratio plot between the reweighted $W$-$p_T$ spectrum and the nominal one, for various entities of the reweighting. The reweighting is performed with a linear function of the $W$-$p_T$. 

129
6.2 Left: example of $\chi^2$ scans for different values of the $W-p_T$ spectrum variation $\alpha$. These examples are obtained with the STK recoil definition, using a linear reweighting. Right: Biases, together with the statistical uncertainties, as a function of the $W-p_T$ spectrum variation $\alpha$. The various colours refer to different recoil definitions. The superimposed straight lines are the linear fit to the points.

6.3 Ranking of the various recoil definitions in terms of statistical uncertainty (x-axis) and systematic uncertainty related to the $W-p_T$ spectrum (y-axis). The latter is expressed in MeV/percent, where percent is the entity of the variation of the $W-p_T$ spectrum variation. The small coloured rectangles represents the statistical uncertainties on the determination of the x and y coordinates of the points due to the finite statistics of the sample, as obtained by splitting the sample in ten sub-samples.

6.4 Comparison of the $Z-p_T$ (left plot) and $Z-p_z$ (right plot) distributions between data and simulation. The $Z$ momentum is reconstructed as the momentum of the dimuon pair. No scale correction on the muon transverse momentum has been applied.

6.5 Sketch of the quantile morphing procedure.

6.6 Data simulation discrepancy on two of the input variables of the regression: magnitude of the TK recoil (left) and number of tracks (right). The first variable is also an $x$ and $y$ variable of the TK recoil definition. These plots show the ratio data/simulation of these variables, in bins of $p_T$ and rapidity of the $Z$ reconstructed boson. The discrepancies are of the order of 10-20%.

6.7 One-dimensional projection of the recoil $p_T$ (left) and of $\Delta \phi_{h,Z}$ (right), and transverse mass distribution computed with the reconstructed muon (right), in $Z+$ events. The first row is obtained with the TK recoil, whereas the second row contains the recoil definition MNPK, as obtained from the regression. The distributions show the data (black), the MC before applying the morphing (red), and the MC after the morphing (blue). As seen in the ratio plots, after the morphing the agreement improves significantly, at a percent level.

6.8 Distributions of the bias as obtained in the self-like fit. Two of the means of these distributions agree with zero within the uncertainty, and the other does within twice the uncertainty. This means that the procedure is not producing any additional bias. Furthermore, the statistical uncertainties on these biases are about 10-20 MeV, confirming that the uncertainty seen in the data-like case are coherent.
6.9 Distributions of the recoil $p_T$ (left) and $\Delta\phi_{\ell,W}$ (right), for the MNPK recoil. The first row, show the distributions in the $Z$ sample (black), $W$ sample (red), and $W_{ZL}$ (blue), whereas the second row shows the opposite. A cut on the $z$ variables ($7 \text{ GeV} < p_T < 11 \text{ GeV}$ and $|p_Z| < 100 \text{ GeV}$) has been imposed in order to mitigate the effect of the difference in the $z$ distribution between $Z$ and $W$. . . . 105
List of Tables

1.1 The six input parameters of the EW precision tests, together with their expressions in terms of the parameters of the SM Lagrangian and of the latest experimental values, as taken from [2].................................................. 9

1.2 Break down of the uncertainties in the $M_W$ prediction within the SM. The Higgs contribution to the $W$ mass is logarithmic, and thus the uncertainty related to this contribution is negligible with respect to the others. Numbers taken from Ref. [1].................................................. 11

1.3 Break down of the uncertainties in the $M_W$ prediction within the SM. The Higgs contribution to the $W$ mass is logarithmic, and thus the uncertainty related to this contribution is negligible with respect to the other. Numbers taken from Ref. [1].................................................. 11

3.1 Break down of the uncertainties in the $W$ mass measurement in CDF and ATLAS analyses. All the numbers are expressed in MeV. Table taken from Ref. [36]........ 37

4.1 Summary of the relative rates and of the features of the $p_T$ spectrum for $W$ bosons, for the different quark-antiquark production channels. The mean, together with its error, is obtained by averaging on the full spectrum. The mean error is very large for the case where the number of events is very small because of the small fraction. MPV stands for the most probable value, i.e. the peak of the distribution. 46

4.2 Summary of the relative rates and of the features of the $p_T$ spectrum for $Z$ bosons, for the different quark-antiquark production channels. The mean, together with its error, is obtained by averaging on the full spectrum. The mean error is very large for the case where the number of events is very small because of the small fraction. MPV stands for the most probable value, i.e. the peak of the distribution. 46
6.1 Summary of the control checks performed in the course of this section. Each check reports the sample on which has been performed, the recoil definition, and the variables used for the templates and for the pseduo-data. The check estimates the bias due to the injected discrepancy between templates and pseudo-data. The uncertainty on the bias is estimated by splitting the sample in 100 sub-samples and repeating the fit for each sub-sample.